

UNIVERSIDADE FEDERAL DO PARANÁ

GUILHERME CORREDATO GUERINO

U2XECs: AVALIAÇÃO DE USABILIDADE E EXPERIÊNCIA DE USUÁRIO DE
SISTEMAS CONVERSACIONAIS

CURITIBA PR

2020

GUILHERME CORREDATO GUERINO

U2XECS: AVALIAÇÃO DE USABILIDADE E EXPERIÊNCIA DE USUÁRIO DE
SISTEMAS CONVERSACIONAIS

Dissertação apresentada como requisito parcial à obtenção do grau de Mestre em Informática no Programa de Pós-Graduação em Informática, Setor de Ciências Exatas, da Universidade Federal do Paraná.

Área de concentração: *Ciência da Computação*.

Orientadora: Natasha Malveira Costa Valentim.

CURITIBA PR

2020

Catálogo na Fonte: Sistema de Bibliotecas, UFPR
Biblioteca de Ciência e Tecnologia

G932u

Guerino, Guilherme Corredato

U2XECS: avaliação de usabilidade e experiência de usuário de sistemas conversacionais [recurso eletrônico] / Guilherme Corredato Guerino. – Curitiba, 2020.

Dissertação - Universidade Federal do Paraná, Setor de Ciências Exatas, Programa de Pós-Graduação em Informática, 2020.

Orientador: Natasha Malveira Costa Valentim

1. Comunicação e tecnologia. 2. Sistemas eletrônicos. 3. Interação homem-computador. 4. Engenharia de Software. I. Universidade Federal do Paraná. II. Valentim, Natasha Malveira Costa. III. Título.

CDD: 004.62

Bibliotecário: Elias Barbosa da Silva CRB-9/1894

TERMO DE APROVAÇÃO

Os membros da Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em INFORMÁTICA da Universidade Federal do Paraná foram convocados para realizar a arguição da Dissertação de Mestrado de **GUILHERME CORREDATO GUERINO** intitulada: **U2XECs: AVALIAÇÃO DE USABILIDADE E EXPERIÊNCIA DE USUÁRIO DE SISTEMAS CONVERSACIONAIS**, sob orientação da Profa. Dra. NATASHA MALVEIRA COSTA VALENTIM, que após terem inquirido o aluno e realizada a avaliação do trabalho, são de parecer pela sua APROVAÇÃO no rito de defesa.

A outorga do título de mestre está sujeita à homologação pelo colegiado, ao atendimento de todas as indicações e correções solicitadas pela banca e ao pleno atendimento das demandas regimentais do Programa de Pós-Graduação.

CURITIBA, 26 de Novembro de 2020.

Assinatura Eletrônica

27/11/2020 08:42:28.0

NATASHA MALVEIRA COSTA VALENTIM

Presidente da Banca Examinadora

Assinatura Eletrônica

01/12/2020 11:36:26.0

DANIELA DE FREITAS GUILHEMINO TRINDADE

Avaliador Externo (UNIVERSIDADE ESTADUAL DO NORTE DO
PARANÁ)

Assinatura Eletrônica

27/11/2020 10:09:52.0

ROBERTO PEREIRA

Avaliador Interno (UNIVERSIDADE FEDERAL DO PARANÁ)

Assinatura Eletrônica

30/11/2020 12:31:24.0

TAYANA UCHOA CONTE

Avaliador Externo (UNIVERSIDADE FEDERAL DO AMAZONAS)

AGRADECIMENTOS

Foram 2 anos e meio de muito esforço, aprendizado, leitura, escrita e dedicação. Gostaria de agradecer a todos que fizeram parte dessa etapa da minha vida e contribuíram de alguma forma para eu chegar até aqui.

Primeiramente, agradeço à minha mãe Andréia e ao meu pai Osmar, por todo o suporte fornecido, pelo amor, compreensão e carinho que vocês têm comigo. Agradeço também por terem me ensinado desde novo que "nada cai do céu", e que para eu alcançar meus objetivos eu preciso correr atrás. De fato, essa lição aprendida logo com 15 anos de idade ecoa na minha cabeça até hoje, e ecoará sempre. Agradeço também minha irmã Lariane e meu cunhado Lucas, por sempre estarem presentes nos momentos de alegria e divertimento.

À Prof^a. Natasha, pela orientação, competência, profissionalismo e dedicação tão importantes. É imensurável o tanto de coisas que aprendi sendo seu orientando nesses anos. Desde assuntos técnicos como Revisões Sistemáticas e Grounded Theory, até percepções profissionais de como ser, de fato, uma boa orientadora. Agradeço também pela confiança no meu trabalho e por todo auxílio.

Aos membros da banca, Prof^a Daniela de Freitas Guilhermino Trindade, Prof^a Tayana Conte e Prof. Roberto Pereira, que tão gentilmente aceitaram participar e colaborar com esta dissertação. Sou fã do trabalho de vocês! É uma honra tê-los comigo nesta etapa.

À minha namorada Milena, por todo amor, carinho, compreensão e apoio nesta caminhada. Obrigado por permanecer ao meu lado, mesmo com a distância e com os dias de loucura com a pesquisa. Obrigado por se fazer presente mesmo de longe, pelo seu sorriso (cheio de dente) e por saber me fazer feliz.

A todos os amigos do Laboratório de Interação Humano-Computador e da UFPR, em especial Leonam, Alisson, Robertino, Ludmilla, Carol, Deógenes e Deivid. Obrigado tanto pelos momentos de aprendizado, quanto os de divertimento.

À CAPES pelo fornecimento da bolsa e fomento da pesquisa e à UFPR pelo apoio financeiro na participação de eventos científicos e pela infraestrutura fornecida.

Por fim, a todos aqueles que contribuíram, direta ou indiretamente, para a realização desta dissertação, o meu sincero agradecimento.

RESUMO

Devido ao aumento do uso das tecnologias nos últimos anos, novas formas de interação estão presentes no cotidiano da sociedade. A interação baseada em voz, característica dos Sistemas Conversacionais (SC), é um exemplo dessas novas formas de interagir. Amazon Alexa, Siri, Google Assistant, Amazon Frame, dispositivos Amazon Echo e Google Home são exemplos de SCs que utilizam a voz do usuário para desempenhar tarefas. Os SCs têm despertado interesse tanto da indústria como da academia, recebendo investimentos e fazendo parte de pesquisa em Interação Humano-Computador (IHC) e Engenharia de Software (ES). Como qualquer outro sistema, é necessário que os SCs forneçam uma boa experiência e que atendam as necessidades de seus usuários. Nesse sentido, a avaliação de Usabilidade e de Experiência do Usuário (*User eXperience* - UX) é vista como etapa importante que contribui com a verificação da qualidade dos SCs. Na avaliação da Usabilidade, geralmente são verificados atributos referentes às metas comportamentais do sistema, como a eficácia, eficiência e satisfação do usuário. Já na avaliação da UX, geralmente são considerados os atributos ligados ao sentimento do usuários, como emoção e motivação. No entanto, através de dois Mapeamentos Sistemáticos da Literatura (MSL), foi identificado que as tecnologias de avaliação utilizadas para avaliar a Usabilidade e/ou UX dos SCs eram genéricas, e poderiam avaliar qualquer tipo software. Além disso, foram identificados alguns questionários de avaliação de Usabilidade ou UX de interfaces conversacionais. Entretanto, estas tecnologias consideram apenas um aspecto de qualidade, Usabilidade, ou UX. Os MSLs também identificaram que alguns pesquisadores utilizam questionários que desenvolveram para seus próprios estudos, sem passar por um processo de avaliação empírica. Sendo assim, o objetivo deste trabalho é fornecer uma tecnologia de avaliação conjunta de Usabilidade e UX específica para SCs, a U2XECS (*Usability and User eXperience Evaluation of Conversational Systems*). A U2XECS é uma tecnologia de avaliação baseada em questionário que fornece afirmações de Usabilidade e UX específicas para avaliar SCs. O objetivo do U2XECS é orientar pesquisadores e desenvolvedores para identificar melhorias e percepções dos usuários nestes sistemas. Além dos MSLs e da proposição da tecnologia, são apresentados também três estudos que foram realizados no processo de elaboração e refinamento da tecnologia: um estudo exploratório, um survey e um estudo de viabilidade. Os resultados evidenciaram pontos positivos da U2XECS relacionados à facilidade de uso, utilidade e intenções de uso. Além disso, foram identificadas oportunidades de melhoria, tais como afirmações ambíguas, mudança na estrutura e no tamanho do questionário.

Palavras-chave: Avaliação de Usabilidade. Avaliação de Experiência de Usuário. Interação Baseada em Voz. Sistemas Conversacionais.

ABSTRACT

Due to the increased use of technologies in recent years, new forms of interaction are present in society. The voice-based interaction, characteristic of Conversational Systems (CSs), is an example of these new interaction forms. Amazon Alexa, Siri, Google Assistant, Amazon Frame, Amazon Echo devices, and Google Home are examples of CSs that use the voice to perform tasks. The CSs have aroused interest from both industry and academia, receiving investments and being part of research in Human-Computer Interaction (HCI) and Software Engineering (SE). Like any other system, CSs must provide a good experience and meet the needs of their users. In this sense, the evaluation of Usability and User eXperience (UX) is seen as an essential step that contributes to verifying the quality of the CSs. In the Usability evaluation, attributes regarding the system's behavioral goals, such as effectiveness, efficiency, and user satisfaction, are usually verified. In the UX evaluation, attributes related to the user's feelings, such as emotion and motivation, are usually considered. However, through two Systematic Mapping Studies (SMS), it was identified that the evaluation technologies used to evaluate the Usability and/or UX of the CSs were generic and could evaluate any software. Besides, some Usability or UX evaluation questionnaires of conversational interfaces were identified. However, these technologies consider only one aspect of quality, Usability, or UX. SMSs also identified that some researchers use questionnaires that they developed for their studies without going through an empirical evaluation process. Therefore, this work aims to provide a CS-specific joint Usability and UX evaluation technology, the U2XECS (Usability and User eXperience Evaluation of Conversational Systems). U2XECS is a questionnaire-based evaluation technology that provides Usability and UX specific statements to evaluate CSs. The goal of U2XECS is to guide researchers and developers to identify improvements and user perceptions in these systems. Besides the SMSs and the technology proposition, three studies that were carried out in the process of elaboration and refinement of the technology are presented: an exploratory study, a survey, and a feasibility study. The results showed positive points of U2XECS related to ease of use, utility, and intentions of use. Besides, opportunities for improvement were identified, such as ambiguous statements, change in the structure and size of the questionnaire.

Keywords: Usability Evaluation. User Experience Evaluation. Voice-Based Interaction. Conversational Systems.

LISTA DE FIGURAS

1.1	Etapas seguidas na metodologia utilizada na pesquisa..	18
2.1	Exemplos de uso dos Sistemas Conversacionais..	20
3.1	Ano de publicação dos artigos selecionados..	33
3.2	Distribuição de artigos por conferência.	34
3.3	Distribuição de artigos por periódico.	35
3.4	Dispositivos de captação utilizados nos estudos.	36
3.5	Aspectos de Usabilidade e UX avaliados pelas tecnologias.	39
3.6	Resultados combinados das subquestões SQ3 e SQ4.	41
3.7	Tecnologias de avaliação utilizadas nos estudos.	42
4.1	Concepção dos participantes em relação ao questionário da U2XECS.	57
5.1	Tecnologias de avaliação utilizadas pelos autores dos artigos retornados no MSL.	67
5.2	Visão temporal do artigos analisados no MSL..	67
5.3	Aspectos de Usabilidade avaliados pelas tecnologias.	70
5.4	Aspectos de UX avaliados pelas tecnologias..	71
5.5	Tipo de coleta de dados utilizado pelas tecnologias.	72
5.6	Categorias dos SCs identificados nos artigos.	75
5.7	Grupos de pessoas que os softwares de voz buscaram auxiliar.	76
7.1	SCs usados pelos respondentes..	86
7.2	Atividades realizadas pelos respondentes.	87
7.3	Frequência de uso dos respondentes.	87
7.4	Uso geral dos respondentes.	88
7.5	Opinião dos respondentes sobre a intuitividade dos SCs.	89
7.6	Opinião dos respondentes sobre a satisfação em relação aos SCs.	89
8.1	Resultados sobre a U2XECS obtidos com as declarações do TAM.	99

LISTA DE TABELAS

3.1	Objetivo do MSL.	24
3.2	Subquestões de pesquisa do MSL.	25
3.3	Bibliotecas de busca utilizadas.	25
3.4	Termos e <i>string</i> de busca utilizados no MSL.	26
3.5	Modelo de tabela para extração de dados.	28
3.6	Total de artigos retornados e selecionados no 1º e 2º filtro.	31
3.7	Resultados do MSL para cada uma das subquestões.	31
3.7	Resultados do MSL para cada uma das subquestões.	32
4.1	Tecnologias que retornaram no MSL que avaliaram uma interação baseada em voz	45
4.1	Tecnologias que retornaram no MSL que avaliaram uma interação baseada em voz	46
4.1	Tecnologias que retornaram no MSL que avaliaram uma interação baseada em voz	47
4.1	Tecnologias que retornaram no MSL que avaliaram uma interação baseada em voz	48
4.2	Aspectos, questões e referências utilizadas na primeira versão da U2XECS. . . .	51
4.2	Aspectos, questões e referências utilizadas na primeira versão da U2XECS. . . .	52
5.1	Objetivo do MSL.	59
5.2	Subquestões de pesquisa do MSL.	60
5.3	Bibliotecas de busca utilizadas.	61
5.4	Termos e <i>string</i> de busca utilizados no MSL.	62
5.5	Tabela usada para a extração de dados	63
5.6	Total de artigos retornados e selecionados no 1º e 2º filtro.	65
5.7	Resultados gerais do segundo MSL.	66
6.1	Alterações realizadas com base no estudo exploratório.	78
6.2	Adição de 18 afirmativas com base nos resultados do MSL.	79
8.1	Resumo da caracterização e identificação de defeitos pelos participantes.	98
A.1	Lista dos artigos extraídos no primeiro MSL.	122
B.1	Lista de tecnologias de avaliação retornadas no primeiro MSL	127

C.1	Lista dos artigos extraídos no segundo MSL	132
D.1	Lista de tecnologias de avaliação retornadas no segundo MSL	136
E.1	Respostas da SQ1, SQ2, SQ2.1, SQ2.2, SQ3 e SQ4.. . . .	141
E.2	Continuação da Tabela E.1	142
E.3	Respostas da SQ2.3.. . . .	143
E.4	Continuação da Tabela E.3.. . . .	144
E.5	Respostas da SQ5.2 e SQ5.3.	145
F.1	Respostas da SQ1, SQ3, SQ4, SQ7 e SQ11.	147
F.2	Continuação Tabela F.1	148
F.3	Aspectos de usabilidade avaliados. Descrição de cada um acima.	149
F.4	Continuação da Tabela F.3.	150
F.5	Continuação da Tabela F.4	151
F.6	Respostas da SQ2.. . . .	152
F.7	Resposta da SQ5.	153
F.8	Respostas da SQ9.. . . .	154
F.9	Respondas da SQ10.	155
F.10	Respostas da SQ12.	155

LISTA DE ACRÔNIMOS

U2X ECS	Usability and User eXperience Evaluation of Conversational System
SC	Sistemas Conversacionais
IHC	Interação Humano-Computador
ES	Engenharia de Software
UX	User eXperience
MSL	Mapeamento Sistemático da Literatura
CS	Conversational System
HCI	Human-Computer Interaction
SE	Software Engineering
SMS	Systematic Mapping Study
IVA	Assistente Virtual Inteligente
SUS	System Usability Scale
UEQ	User Experience Questionnaire
NUI	Natural User Interface
GQM	Goal-Question-Metric
PICOC	Population, Intervention, Comparison, Outcome e Context
SQ	Subquestão
SDK	Software Development Kit
GEQ	Game Experience Questionnaire
SASSI	Subjective Assessment of Speech System Interfaces
TCLE	Termo de Consentimento Livre e Esclarecido
TAM	Technology Acceptance Model
FU	Facilidade de uso percebida
UP	Utilidade percebida
IU	Intenção de uso futuro
QA	Questões abertas
GT	Grounded Theory
SJR	Scimago Journal Rankings
CORE	The Computing Research and Education
ACE	Automated-Caption Evaluation
WER	Word Error Rate
HPV	Human Papilloma Virus

SUMÁRIO

1	INTRODUÇÃO	14
1.1	MOTIVAÇÃO	14
1.2	PROBLEMA	15
1.3	OBJETIVOS	16
1.3.1	Objetivos específicos:	16
1.4	METODOLOGIA	16
1.5	ORGANIZAÇÃO DA DISSERTAÇÃO	18
2	BACKGROUND TEÓRICO E TRABALHOS RELACIONADOS	20
2.1	SISTEMAS CONVERSACIONAIS	20
2.2	USABILIDADE E EXPERIÊNCIA DE USUÁRIO	21
2.3	TRABALHOS RELACIONADOS	22
3	MAPEAMENTO SISTEMÁTICO DA LITERATURA SOBRE TECNOLOGIAS DE AVALIAÇÃO DE USABILIDADE E/OU UX DE NUI	24
3.1	PROTOCOLO DO MAPEAMENTO SISTEMÁTICO DA LITERATURA	24
3.1.1	Objetivo	24
3.1.2	Questões de pesquisa	24
3.1.3	Estratégia utilizada para pesquisa da publicações	25
3.1.4	Critérios de Seleção de Artigos	27
3.1.5	Processo de Seleção de Artigos	27
3.1.6	Definição de estratégia de extração de dados	28
3.1.7	Artigos selecionados após a condução do MSL	30
3.2	RESULTADOS OBTIDOS	31
3.2.1	Visão geral dos resultados obtidos	31
3.2.2	Ano de publicação	33
3.2.3	Locais de publicação	34
3.2.4	Tipo de interação da NUI avaliada (SQ1)	35
3.2.5	Dispositivo para a captação de interações naturais	36
3.2.6	Critério de qualidade da tecnologia de avaliação (SQ2)	37
3.2.7	Tecnologias de avaliação específicas ou não (SQ3)	38

3.2.8	Uso de tecnologias baseadas em outras já existentes ou não (SQ4)	40
3.2.9	Avaliação empírica (SQ5).	40
3.3	SÍNTESE DO CAPÍTULO	43
4	PROPOSTA DA PRIMEIRA VERSÃO DA U2XECS	44
4.1	INTRODUÇÃO	44
4.2	ANÁLISE DAS TECNOLOGIAS DE AVALIAÇÃO IDENTIFICADAS NO 1º MSL.	45
4.3	PRIMEIRA VERSÃO DA U2XECS	50
4.4	TEMPLATE DO QUESTIONÁRIO DA PRIMEIRA VERSÃO DA U2XECS . .	53
4.5	ESTUDO EXPLORATÓRIO COM A PRIMEIRA VERSÃO DA U2XECS . . .	54
4.5.1	Planejamento	54
4.5.2	Execução	55
4.5.3	Análise	56
5	MAPEAMENTO SISTEMÁTICO DA LITERATURA SOBRE TECNOLO- GIAS DE AVALIAÇÃO DE USABILIDADE E/OU UX PARA SCS.	59
5.1	PROTOCOLO DO MAPEAMENTO SISTEMÁTICO DA LITERATURA. . . .	59
5.1.1	Objetivo	59
5.1.2	Questões de pesquisa	59
5.1.3	Estratégia utilizada para pesquisa das publicações	60
5.1.4	Critérios de seleção de artigos	61
5.1.5	Processo de seleção de artigos	62
5.1.6	Definição da estratégia de extração de dados	63
5.1.7	Artigos selecionados após a condução do MSL	65
5.2	RESULTADOS OBTIDOS	65
5.2.1	Visão geral dos resultados obtidos	65
5.2.2	Ano de publicação	67
5.2.3	Locais de publicação	68
5.2.4	SQ1. Critério de Qualidade.	68
5.2.5	SQ2. Aspectos de Usabilidade e UX	69
5.2.6	SQ3. Especificidade da tecnologia de avaliação	69
5.2.7	SQ4. Base da tecnologia de avaliação	71
5.2.8	SQ5. Método de coleta das respostas dos participantes	72

5.2.9	SQ6. Características das tecnologias de avaliação	73
5.2.10	SQ7. Tipo de análise	73
5.2.11	SQ8. Função do SC	74
5.2.12	SQ9. Categoria do SC	75
5.2.13	SQ10. Grupo de usuários de SC	75
5.2.14	SQ11. Avaliação empírica da tecnologia de avaliação	76
5.3	SÍNTESE DO CAPÍTULO	77
6	PROPOSTA DA SEGUNDA VERSÃO DA U2XECS	78
6.1	ALTERAÇÕES REALIZADAS	78
6.2	TEMPLATE DO QUESTIONÁRIO DA SEGUNDA VERSÃO DA U2XECS . .	80
7	SURVEY EXPLORATÓRIO	84
7.1	PLANEJAMENTO E EXECUÇÃO	84
7.2	RESULTADOS E DISCUSSÃO	85
7.2.1	Análise Quantitativa	85
7.2.2	Análise Qualitativa	89
7.3	SÍNTESE DO CAPÍTULO	93
8	ESTUDO DE VIABILIDADE	94
8.1	PLANEJAMENTO	94
8.1.1	Contexto.	94
8.1.2	Seleção dos Participantes	94
8.1.3	Seleção de Indicadores	95
8.1.4	Instrumentação	95
8.1.5	Preparação.	96
8.2	EXECUÇÃO	96
8.3	ANÁLISE QUANTITATIVA	97
8.4	ANÁLISE DA ACEITAÇÃO DA U2XECS	99
8.4.1	Facilidade de uso percebida.	99
8.4.2	Utilidade percebida	99
8.4.3	Intenção de uso	100
8.5	ANÁLISE QUALITATIVA	100
8.5.1	Benefícios do questionário	101

8.5.2	Pontos negativos	101
8.5.3	Sugestões de melhoria	101
8.6	SÍNTESE DO CAPÍTULO	102
9	PROPOSTA DA TERCEIRA VERSÃO DA U2X ECS	103
9.1	ALTERAÇÕES	103
9.2	TEMPLATE DO QUESTIONÁRIO DA TERCEIRA VERSÃO DA U2X ECS . .	105
10	CONSIDERAÇÕES FINAIS	108
10.1	CONTRIBUIÇÕES	109
10.2	PERSPECTIVAS PARA TRABALHOS FUTUROS.	109
10.3	PUBLICAÇÕES	110
	REFERÊNCIAS	111
	APÊNDICE A – LISTA DOS ARTIGOS EXTRAÍDOS NO PRIMEIRO	
	MSL.	122
	APÊNDICE B – LISTA DE TECNOLOGIAS DE AVALIAÇÃO RETOR-	
	NADAS NO PRIMEIRO MSL	127
	APÊNDICE C – LISTA DOS ARTIGOS EXTRAÍDOS NO SEGUNDO	
	MSL.	132
	APÊNDICE D – LISTA DE TECNOLOGIAS DE AVALIAÇÃO RETOR-	
	NADAS NO SEGUNDO MSL	136
	APÊNDICE E – RESUMO 1º MSL	141
	APÊNDICE F – RESUMO 2º MSL	146

1 INTRODUÇÃO

Este capítulo apresenta a motivação dessa pesquisa, assim como os questionamentos levantados para a sua elaboração, além dos objetivos propostos e a metodologia aplicada durante o processo de investigação e análise científica, e por fim, a organização dessa dissertação de mestrado.

1.1 MOTIVAÇÃO

Os Sistemas Conversacionais (SCs) estão recebendo cada vez mais investimentos na indústria e sendo discutidos pela pesquisa de Interação Humano-Computador (IHC) e Engenharia de Software (ES). Do lado da indústria, estima-se que somente o mercado de Assistentes Virtuais Inteligentes (IVAs) alcançará 45,1 bilhões de dólares em 2027 (Grand View Research, Inc, 2020). Do lado da pesquisa, as áreas de IHC e ES têm fornecido várias contribuições e estudos com diversos contextos e perfis de usuários, como usuários idosos (Wulf et al., 2014; Trajkova e Martin-Hammond, 2020) e crianças (Lovato e Piper, 2015), bem como estudos de caso (Candello et al., 2019; Jeong et al., 2019) e revisões de literatura (Kocaballi et al., 2020; Laranjo et al., 2018). O grande interesse neste tipo de sistema é porque ele é amplamente utilizado em aplicações para smartphones (por exemplo, Siri e Alexa), dispositivos físicos para automação doméstica (por exemplo, Amazon Echo Dot e Google Home), integração com sistemas para carros (integração Google Assistant) e até mesmo óculos inteligentes (Amazon Echo Frame).

Neste sentido, o SC deve alcançar um nível de qualidade para ser realmente útil, fácil de usar e agradável aos seus usuários. Uma forma de verificar a qualidade desses sistemas, é por meio da avaliação de Usabilidade e Experiência de Usuário (*User eXperience - UX*) (Madan e Dubey, 2012). Uma avaliação da Usabilidade e UX permite aos desenvolvedores e pesquisadores ter uma percepção - positiva ou negativa - do sistema desenvolvido antes de torná-lo disponível para a sociedade. A partir desta etapa e dos resultados obtidos, os pontos negativos podem ser melhorados.

Kocaballi et al. (2018) revelam que existem várias interpretações de Usabilidade e UX para SCs, uma vez que o UX é usado: (i) como Usabilidade; (ii) como algo além da Usabilidade; ou (iii) como satisfação do usuário. Neste trabalho, é interpretada a Usabilidade de acordo com a ISO 9241-210 (2019), sendo "a medida em que um sistema, produto ou serviço pode ser usado por usuários especificados para atingir objetivos específicos com eficácia, eficiência e satisfação em um contexto de uso especificado". A UX também foi interpretada de acordo com a ISO 9241-210 (2019), sendo "as percepções e respostas do usuário que resultam do uso e/ou do uso antecipado de um sistema, produto ou serviço". Portanto, enquanto a avaliação de Usabilidade verifica metas comportamentais (como eficiência e eficácia), a avaliação de UX compreende

os conceitos relacionados às percepções subjetivas dos usuários do sistema (como motivação, prazer e emoção).

Tanto nos métodos avaliativos de Usabilidade quanto nos de UX, existem diversos tipos de tecnologias (modelos, *frameworks*, abordagens, questionários, entre outros (Santos et al., 2012)) que podem ser utilizadas. Um exemplo é o *System Usability Scale* (SUS) (Brooke, 1996), utilizado para avaliar a Usabilidade (mais especificamente a satisfação do usuário) e o *User Experience Questionnaire* (UEQ) (Laugwitz et al., 2008), utilizado para avaliar a UX de acordo com vários aspectos como motivação e o quão interessante foi utilizar aquele produto. No entanto, as tecnologias de avaliação existentes geralmente são genéricas, ou seja, não focam em um tipo específico de software. Por um lado, isso é uma vantagem pois aumenta a aplicabilidade da tecnologia de avaliação. Porém, em alguns casos, é necessário direcionar melhor a avaliação para que ela leve em consideração aspectos específicos do software que está sendo avaliado e identifique potenciais problemas que os usuários possam vir a experimentar.

Sendo assim, foi vista a necessidade de investigar como avaliar a Usabilidade e UX direcionadas aos SCs, levando em consideração uma interação por voz e "não convencional".

1.2 PROBLEMA

Apesar da popularidade, os SCs ainda possuem algumas falhas durante o uso, e a aceitação do usuário é baixa (Moore, 2017). Alguns estudos sugerem vários fatores que contribuem para a frustração dos usuários, tais como questões de confiança (Wulf et al., 2014), habitabilidade (o equilíbrio no design entre a linguagem que as pessoas usam e a linguagem que um SC pode aceitar) (Kocaballi et al., 2020; Moore, 2017), e privacidade (Moore et al., 2016). Por conta disso, é necessário que se atente a este tipo de sistema, visto que tecnologias genéricas podem influenciar as avaliações de interações que não são tradicionalmente usadas, devido ao fato da baixa familiaridade do usuário com a interação (Blake, 2011). Quiñones *et al.* (2018) mostram que há uma preocupação dos pesquisadores em geral de avaliar contexto específicos de aplicações e suas características. Portanto, essas avaliações devem ser evidenciadas e realizadas a partir de uma metodologia bem definida.

Além disso, pesquisas atuais buscam descobrir e justificar se a Usabilidade e a UX são conceitos trabalhados em conjunto e estabelecem dependência, ou se são conceitos que devem ser abordados separadamente. Segundo Nielsen (1993), a Usabilidade é um componente de aceitabilidade do sistema, que visa verificar se a aplicação é boa o suficiente para satisfazer todas as exigências dos usuários. Já a UX, foca no design e desenvolvimento dos produtos com o objetivo de melhorar a percepção e sentimento do usuário ao utilizá-los (Hassenzahl e Tractinsky, 2006).

Nesta pesquisa, considerou-se que ambos os conceitos são importantes para a qualidade do SC. Foi levado em consideração também que, nessa concepção, são conceitos separados mas que devem ser considerados em conjunto durante a avaliação. A avaliação conjunta de

Usabilidade e UX permite que sejam verificados, simultaneamente, os aspectos ligados às metas comportamentais e os ligados aos sentimentos dos usuários (Väättäjä et al., 2009). No entanto, não se identificou na literatura uma tecnologia que avalie tanto a Usabilidade quanto a UX especificamente para um SC.

Sendo assim, foi vista a necessidade de propor uma tecnologia de avaliação de Usabilidade e UX exclusiva para os SCs, com o objetivo de avaliá-lo inteiramente, considerando suas especificidades. Portanto, a questão que direcionou esta pesquisa foi: *“Como avaliar a Usabilidade e a Experiência de Usuário de um Sistema Conversacional em conjunto?”*

1.3 OBJETIVOS

Este trabalho tem como principal objetivo propor uma tecnologia de avaliação de Usabilidade e UX para SCs. Para isso, tanto o conceito de Usabilidade quanto o de UX serão levados em consideração na mesma tecnologia de avaliação, de modo a potencializar a satisfação do usuário com este tipo de software.

1.3.1 Objetivos específicos:

1. Construir um corpo de conhecimento das tecnologias que estão sendo utilizadas para avaliar a Usabilidade e/ou UX dos softwares que utilizam formas de interação não convencional e SCs;
2. Elaborar uma tecnologia de avaliação de Usabilidade e UX específica para o contexto de SCs;
3. Avaliar e evoluir a tecnologia proposta por meio de estudos, apresentando evidências empíricas do desempenho da tecnologia no contexto de SCs.

1.4 METODOLOGIA

Para atingir os objetivos definidos nesta pesquisa, foi utilizada uma metodologia baseada em evidências, que se beneficia de estudos primários e secundários (Mafra e Travassos, 2006). Os estudos primários nos permitem testar hipóteses, enquanto os secundários coletar dados importantes sobre um tema específico na literatura.

Os estudos secundários desta pesquisa foram realizados por meio de dois Mapeamentos Sistemáticos da Literatura (MSL), com o objetivo de identificar soluções existentes para o problema de pesquisa e desafios relacionados. Este conhecimento obtido através de estudos secundários, tem por finalidade criar um corpo de conhecimento e evidências como alicerce para a definição da proposta desta pesquisa. Segundo Kitchenham e Charters (2007), o MSL pode ser definido como uma forma de identificar, avaliar e interpretar todas as pesquisas disponíveis que são relevantes para uma particular questão de pesquisa, área ou fenômeno de interesse.

A proposta, então, foi avaliada por meio de estudos primários (estudos experimentais). Estudos primários permitem que os pesquisadores respondam perguntas, examinem ou identifiquem problemas em novas propostas (Shull et al., 2001).

A Figura 1.1 apresenta a metodologia deste trabalho, considerando estudos secundários para coletar informações sobre as tecnologias que avaliam a Usabilidade e/ou a UX de SCs. Além disso, beneficia-se de estudos primários para avaliar a proposta da tecnologia e obter feedback para evolução do estudo. As etapas da metodologia desta pesquisa são apresentadas a seguir:

1. Mapeamento Sistemático da Literatura para Interfaces Naturais de Usuário (*Natural User Interfaces* - NUI): esta etapa consistiu na realização de um estudo secundário para coletar informações sobre tecnologias que avaliam a Usabilidade e/ou a UX em softwares que utilizam interações naturais, onde a interação por voz está inserida.
2. Proposta da Primeira Versão da *Usability and User eXperience Evaluation of Conversational Systems* (U2XECS): nesta etapa, propõe-se a primeira versão da tecnologia U2XECS. Para isso, foi selecionado um conjunto de tecnologias de avaliação de Usabilidade e/ou UX identificadas a partir da etapa (1), com o objetivo de verificar os pontos fracos e os pontos fortes a fim de reunir um conjunto de recursos a serem incorporados na proposta. Os dados foram condensados para se obter uma tecnologia que atenda aos aspectos da Usabilidade e UX em conjunto.
3. Estudo Exploratório: o estudo exploratório foi realizado com o objetivo de verificar a facilidade de uso, utilidade percebida e as intenções de uso da primeira versão da U2XECS. Buscou-se também, por meio deste estudo, captar possíveis orientações fornecidas pelos participantes a fim de refinar e evoluir a tecnologia.
4. Mapeamento Sistemático da Literatura para Software de Interação Baseada em Voz: esta etapa consistiu na realização de um segundo estudo secundário para coletar informações sobre tecnologias que avaliam a Usabilidade e/ou a UX especificamente em softwares de interação baseada em voz.
5. Proposta da Segunda Versão da U2XECS: nesta etapa, propõe-se uma evolução da U2XECS. Para esta evolução, foram considerados os resultados do estudo exploratório (3) e do segundo MSL (4).
6. *Survey* Exploratório: nesta etapa, foi realizado um *survey* com usuários e ex-usuários de sistemas conversacionais. O objetivo foi obter informações de usuários finais para identificar novos *insights* a serem incorporados na U2XECS.
7. Estudo de Viabilidade: o estudo de viabilidade foi realizado para verificar a eficiência dos participantes ao utilizar a U2XECS na identificação de defeitos em um SC a partir de aspectos de Usabilidade e UX. Além disso, foi verificado também a opinião

dos participantes em relação a facilidade de uso, utilidade, intenções de uso futuro e melhorias para a U2XECS.

8. Terceira Versão da U2XECS: a partir dos resultados do *survey* (6) e do estudo de viabilidade (7), propõe-se uma evolução da U2XECS para sua terceira (e atual) versão.

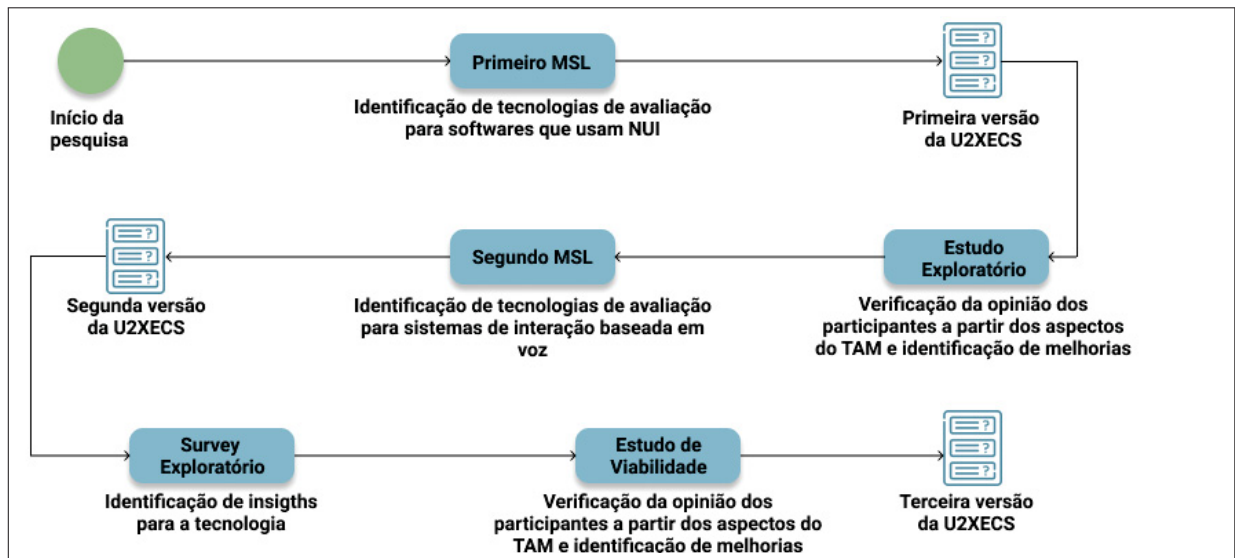


Figura 1.1: Etapas seguidas na metodologia utilizada na pesquisa.

1.5 ORGANIZAÇÃO DA DISSERTAÇÃO

Esta dissertação está organizada em outros nove capítulos, além deste primeiro capítulo de introdução, que apresentou a motivação, o problema, os objetivos e a metodologia. A organização do texto deste trabalho segue a estrutura abaixo:

- **Capítulo 2 - *Background Teórico e Trabalhos Relacionados***: descreve definições de Sistemas Conversacionais, Usabilidade e UX que baseiam esta pesquisa. Além disso, mostra trabalhos relacionados com os temas abordados neste trabalho.
- **Capítulo 3 - *Primeiro Mapeamento Sistemático da Literatura***: mostra o planejamento, execução e resultados de um estudo secundário realizado com o propósito de identificar as tecnologias de avaliação de Usabilidade e UX que têm sido usadas para avaliar os softwares que possuem uma NUI;
- **Capítulo 4 - *Proposta da primeira versão da U2XECS***: descreve a formulação da proposta U2XECS, suas afirmativas e *template*. Além disso, mostra a execução de um estudo exploratório que objetivou verificar a aceitação da U2XECS e identificar melhorias para a tecnologia;

- **Capítulo 5 - Segundo Mapeamento Sistemático da Literatura:** mostra o planejamento, execução e resultados de um estudo secundário realizado com o propósito de identificar as tecnologias de avaliação de Usabilidade e UX que têm sido usadas para avaliar os software de interação por voz;
- **Capítulo 6 - Proposta da segunda versão da U2XECS:** descreve as alterações realizadas no processo evolutivo e o *template* da segunda versão da U2XECS;
- **Capítulo 7 - Survey Exploratório:** mostra o planejamento, execução, resultados e discussão de um *survey* realizado com usuários e ex-usuários de Sistemas Conversacionais para verificar o uso e dificuldades desse tipo de sistema;
- **Capítulo 8 - Estudo de Viabilidade:** mostra o planejamento, execução, resultados e discussão de um estudo realizado com a U2XECS para verificar a viabilidade da tecnologia e identificar melhorias;
- **Capítulo 9 - Proposta da terceira versão da U2XECS:** descreve as alterações realizadas no processo evolutivo e o *template* da terceira (e atual) versão;
- **Capítulo 10 - Considerações Finais:** mostra as conclusões do trabalho, as contribuições, perspectivas futuras e as publicações científicas obtidas.

2 BACKGROUND TEÓRICO E TRABALHOS RELACIONADOS

Este capítulo apresenta conceitos e definições de Sistemas Conversacionais (SCs), Usabilidade e Experiência de Usuário (UX) que fundamentaram essa pesquisa. Além disso, este capítulo apresenta alguns trabalhos relacionados a esta pesquisa.

2.1 SISTEMAS CONVERSACIONAIS

Os Sistemas Conversacionais (SCs) são softwares que utilizam voz ou outra modalidade para realizar alguma interação de conversação (McTear et al., 2016). O uso de SCs tem aumentado nos últimos anos, sendo usados como aplicativos móveis, chatbots, dispositivos físicos e Assistentes Virtuais Inteligentes (AVIs). A Figura 2.1 apresenta alguns exemplos de SCs: (a) Dispositivo Amazon Echo, baseado na assistente virtual Alexa; (b) Dispositivo Google Home, baseado no Google Assistant; (c) Assistente virtual Cortana (Windows); (d) Assistente Virtual Siri (iOS). Os SCs geralmente são usados pela sociedade para acessar informações, entretenimento, enviar mensagens e controlar outros dispositivos inteligentes.

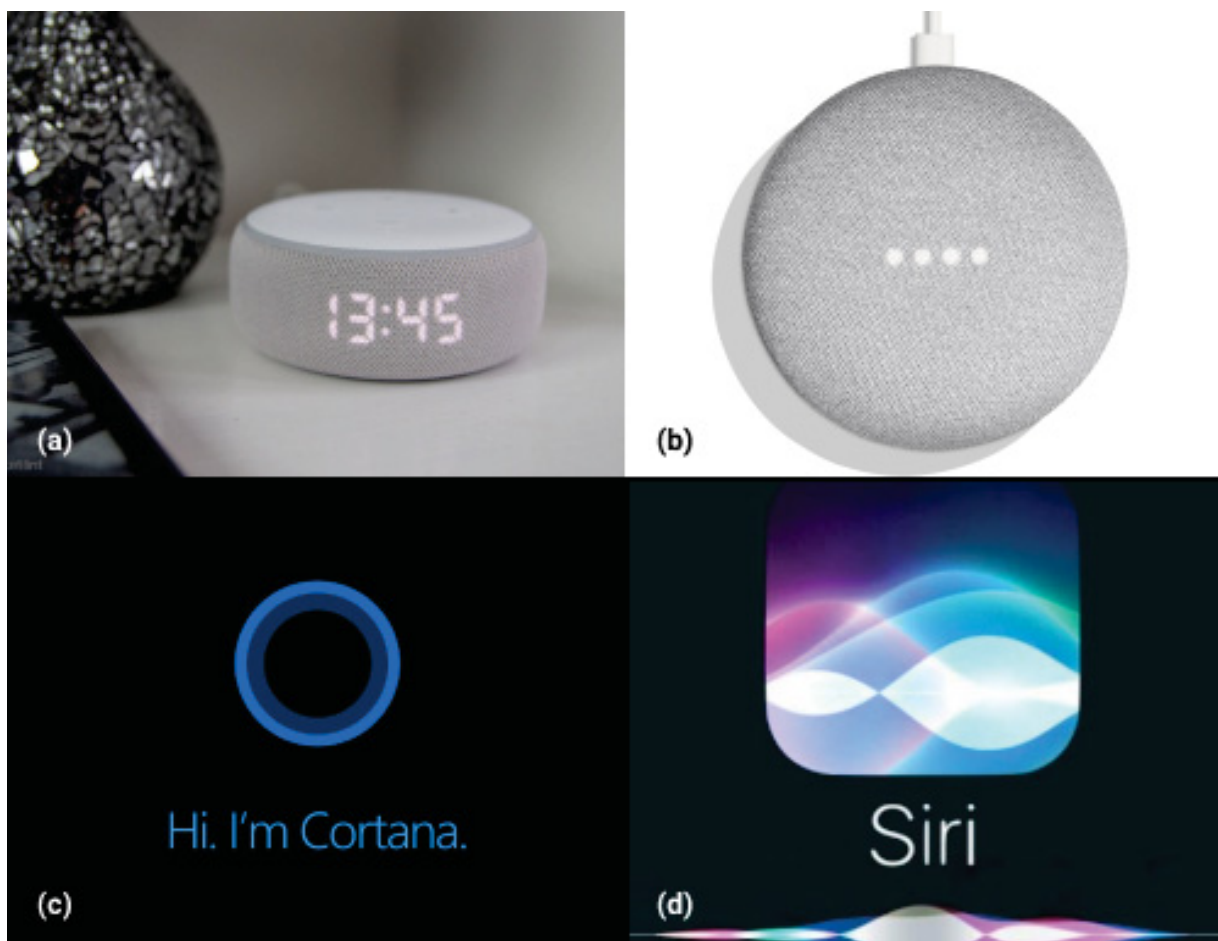


Figura 2.1: Exemplos de uso dos Sistemas Conversacionais.

Do lado da indústria, estima-se que com o alto uso de smartphones e a expansão da Inteligência Artificial, somente o mercado de AVIs chegará a 4,6 bilhões de dólares até o final de 2022 (Kamitis, 2016). Uma pesquisa mais recente mostrou que o mercado de AVIs atingirá 45,1 bilhões de dólares em 2027 (Grand View Research, Inc, 2020).

O alto uso e os investimentos recebidos também despertaram interesse da comunidade de Interação Humano-Computador (IHC) (Porcheron et al., 2018). Os estudos nos últimos anos têm se concentrado em entender como melhorar a conversação entre usuários e sistemas, de modo a fornecer uma experiência relevante para quem usa (Spillane et al., 2017).

Dessa forma, os experimentos e pesquisas com SCs têm sido difundidos em diversas áreas, como saúde (Bickmore et al., 2010), cuidado com idosos (Wulf et al., 2014; Trajkova e Martin-Hammond, 2020; Vardoulakis et al., 2012), crianças (Lovato e Piper, 2015) e educação (Saerbeck et al., 2010). Além disso, podem ser identificados também estudos secundários abordando o tema (Kocaballi et al., 2020; Laranjo et al., 2018). Estes trabalhos serão discutidos na Seção 2.3.

Como qualquer outro tipo de sistema, e movidos por uma demanda social para seu uso, os SCs precisam atingir um nível de qualidade para serem verdadeiramente úteis aos seus usuários. Moore (2017) mostra que os SCs ainda apresentam algumas falhas durante o uso, o que pode comprometer a sua aceitação.

Nesse sentido, as avaliações de Usabilidade e UX podem ser úteis na identificação de problemas com estes sistemas antes de inseri-los na sociedade. A partir destas avaliações, os pesquisadores e desenvolvedores podem identificar pontos negativos no SC a serem melhorados.

2.2 USABILIDADE E EXPERIÊNCIA DE USUÁRIO

A área de IHC está em constante mudança devido à evolução tecnológica e os novos desafios advindos da sociedade em geral (Fuchsberger et al., 2012). Para Bødker (2006), essa evolução pode ser dividida em três grandes “ondas”. A primeira onda foi caracterizada pelas diretrizes, testes e métodos formais para tornar as tecnologias mais utilizáveis. A segunda onda, apresentou o *design* centrado no usuário, com conceitos como *design* participativo e pesquisa contextual. A terceira onda aborda ainda mais o foco do usuário, apresentando conceitos referentes às emoções. Desta forma, é possível perceber uma evolução dos termos da primeira para terceira onda, da Usabilidade à UX, mas sem que um ou outro perca o seu valor e características. Sendo assim, mostra-se necessário que as tecnologias avaliativas “englobem as ondas” e se adequem ao que se está emergindo na sociedade. Ou seja, é preciso abordar associadamente os aspectos de Usabilidade, opiniões e emoções dos usuários (UX) nas tecnologias avaliativas para que elas acompanhem a evolução e as mudanças tecnológicas e sociais.

Apesar de autores divergirem opiniões sobre as dependências entre os conceitos de Usabilidade e UX, Hassenzahl (2004) mostra que dois tipos de atributos são percebidos pelos usuários ao utilizar um sistema, a qualidade pragmática e a qualidade hedônica. Os atributos

pragmáticos estão conectados à necessidade dos usuários de atingir metas comportamentais (Hassenzahl, 2004). De acordo com este autor, a realização de uma tarefa com eficiência e eficácia pode ser entendida como uma qualidade pragmática de um produto. Já a qualidade hedônica está ligada ao “eu” do usuário (Hassenzahl, 2004). Para o autor, os atributos desse tipo de qualidade fornecem estímulo para o caráter desafiador e inovador dos usuários, comunicando com valores pessoais e envolvendo as suas experiências.

Apesar dos aspectos da Usabilidade e UX terem sido bastante popularizados nos últimos anos, é importante ressaltar que não são muitos os trabalhos identificados na literatura que envolvem SCs e que avaliam esses dois conceitos de IHC em conjunto, a Usabilidade e a UX. A avaliação conjunta da Usabilidade e da UX permite que os pesquisadores/desenvolvedores verifiquem simultaneamente as metas comportamentais do sistemas e os aspectos ligados aos sentimentos do usuário (Väätäjä et al., 2009), economizando custo e esforço, e otimizando o tempo de avaliação .

2.3 TRABALHOS RELACIONADOS

Alguns estudos secundários abordaram o tema de SCs de uma forma geral. Em Laranjo et al. (2018), os autores investigaram características, aplicações e medidas de avaliação de SCs para o cuidado da saúde. Os autores identificaram que o uso de SCs em saúde é um campo emergente de pesquisa, e poucos trabalhos avaliaram sua eficácia. Apesar de ser um estudo de medidas de avaliação em um contexto conversacional, os autores concentraram suas pesquisas na área da saúde.

Ren et al. (2019) realizaram um MSL para caracterizar as tecnologias de avaliação de Usabilidade utilizadas em chatbots. Os resultados mostraram que a Usabilidade de chatbot é um campo em desenvolvimento, com estudos experimentais muito informais. Apesar do estudo ter sido específico para o contexto da voz, o trabalho se concentrou apenas em chatbots e Usabilidade.

Kocaballi et al. (2018) realizaram um estudo comparativo de seis questionários que avaliam a UX de SCs: AttrakDiff, Subjective Assessment of Speech System Interfaces (SASSI), Speech User Interface Service Quality (SUISQ), Mean Opinion Scale Expanded (MOS-X), Paradise e System Usability Scale (SUS). Os resultados mostraram que nenhum dos questionários avaliou suficientemente algumas dimensões da UX, tais como encantamento, diversão e motivação. Embora o estudo forneça resultados importantes para o estado da arte, o trabalho focalizou apenas o contexto UX e questionários. Além disso, a seleção dos seis questionários foi feita com base na experiência dos próprios pesquisadores.

Com a popularidade de SCs como Amazon Echo e Google Home, a comunidade de IHC desenvolveu um interesse considerável neste tópico (Porcheron et al., 2018). Alguns estudos verificam o uso de chatbots, dispositivos físicos e AVIs. Por exemplo, em perfis de usuários específicos, a Siri foi avaliada com usuários idosos (Wulf et al., 2014) e crianças (Lovato e Piper,

2015). Os resultados revelaram que várias dificuldades foram encontradas, como a qualidade do reconhecimento da fala e problemas com a conexão da internet.

A aceitação de chatbots como SCs foi avaliada por estudantes das Universidades do Reino Unido (Almahri et al., 2020). Os resultados mostraram que a expectativa de desempenho, a expectativa de esforço e o hábito são os principais fatores da intenção comportamental dos estudantes de usar chatbots. Além disso, a influência do público durante o uso de chatbots no espaço físico também foi verificada (Candello et al., 2019). Os resultados sugerem que os projetistas desenvolvam SCs com base na observação ou não da interação do usuário por outras pessoas.

A Alexa também foi avaliada com usuários idosos (Trajkova e Martin-Hammond, 2020). Os resultados demonstraram que vários participantes abandonaram o uso da Alexa porque não identificaram sua utilidade. As limitações dos AVIs também foram abordadas com usuários pouco frequentes (Cowan et al., 2017). Os resultados mostraram pontos que corroboram com as opiniões dos usuários especializados e outros novos pontos identificados pelos usuários pouco frequentes.

Como mencionado, alguns estudos secundários sobre avaliação de Usabilidade e UX foram realizados. Entretanto, não foram encontrados estudos secundários que combinassem estes dois conceitos, avaliação de Usabilidade e UX, e SCs em geral. Para investigar o cenário de avaliações de interações não-convencionais de uma forma abrangente antes de focalizar em SCs, foi realizado um Mapeamento Sistemático da Literatura com foco em NUIs, descrito na seção a seguir.

3 MAPEAMENTO SISTEMÁTICO DA LITERATURA SOBRE TECNOLOGIAS DE AVALIAÇÃO DE USABILIDADE E/OU UX DE NUI

Este capítulo apresenta um MSL das tecnologias de avaliação de Usabilidade e/ou UX usadas em sistemas que implementam NUIs, abordando a seguinte questão de pesquisa: “*Quais tecnologias são usadas para avaliar a Usabilidade e/ou a UX de um software que implementa as Interfaces Naturais de Usuário?*”. Também é mostrado como o MSL foi planejado e conduzido, além dos resultados e lacunas de pesquisa obtidos por meio dele.

3.1 PROTOCOLO DO MAPEAMENTO SISTEMÁTICO DA LITERATURA

De acordo com a justificativa mencionada, foi observada a necessidade de se realizar um Mapeamento Sistemático da Literatura (MSL) para identificar e compreender o que já foi realizado sobre avaliação de Usabilidade e UX (UX) em softwares que implementam as Interfaces Naturais de Usuário (NUI), e as lacunas que essa vertente de pesquisa possui para serem exploradas.

3.1.1 Objetivo

A descrição do objetivo da MSL foi baseada no paradigma *Goal-Question-Metric* (GQM) (Basili e Rombach, 1988), e está descrita na Tabela 3.1.

Tabela 3.1: Objetivo do MSL.

Analisar	publicações científicas
Com o propósito de	caracterizar
Em relação a	tecnologias que avaliam a Usabilidade e/ou a UX em softwares que implementam as NUIs
Do ponto de vista dos	pesquisadores de IHC
No contexto de	publicações disponíveis em SCOPUS, ACM, IEEEExplore, Engineering Village e Science Direct.

3.1.2 Questões de pesquisa

A questão de pesquisa principal deste mapeamento é: “*Quais tecnologias são usadas para avaliar a Usabilidade e/ou a UX de um software que implementa as NUIs?*” Além desta questão de pesquisa, foram definidas subquestões para responder questionamentos específicos sobre cada tecnologia de avaliação. As subquestões são apresentadas na Tabela 3.2.

Tabela 3.2: Subquestões de pesquisa do MSL.

Subquestões de Pesquisa
SQ1. A tecnologia avalia qual tipo de interação da NUI?
SQ2. Qual o critério de qualidade da tecnologia de avaliação?
SQ2.1 Quais os tipos de tecnologia de avaliação de Usabilidade foram usados?
SQ2.2 Quais os tipos de tecnologia de avaliação de UX foram usados?
SQ2.3 Quais aspectos de Usabilidade e/ou UX a tecnologia avalia?
SQ3. A tecnologia é específica para software que implementa NUI ou é para software em geral?
SQ4. A avaliação utiliza uma tecnologia de avaliação já existente? Se sim, qual?
SQ5. O software que implementa a NUI foi avaliado experimentalmente?
SQ5.1 Como e por qual tecnologia de avaliação foi realizada a avaliação empírica?
SQ5.2 A avaliação empírica foi feita na academia, indústria ou laboratório?
SQ5.3 A avaliação empírica foi analisada de maneira quantitativa ou qualitativa?

3.1.3 Estratégia utilizada para pesquisa da publicações

A estratégia de busca deve propiciar a integridade da pesquisa (Kitchenham e Charters, 2007). Para isso, o MSL deve ser realizado conforme uma estratégia de pesquisa predefinida. A estratégia é descrita a seguir.

- **Escopo da pesquisa:** a pesquisa foi feita a partir de bibliotecas digitais através de mecanismo de busca avançada utilizando string de busca. Na Tabela 3.3 estão algumas informações sobre as bibliotecas digitais utilizadas neste MSL, como nome e link.

Tabela 3.3: Bibliotecas de busca utilizadas.

Nome da fonte	Link
Scopus	http://www.scopus.com/home.url
IEEEExplore	https://ieeexplore.ieee.org/Xplore/home.jsp
ACM Digital Library	http://portal.acm.org/dl.cfm
Engineering Village	https://www.engineeringvillage.com/search/quick.url
Science Direct	https://www.sciencedirect.com/

Estas bibliotecas foram escolhidas: (1) por fornecerem um eficiente mecanismo de busca, (2) por permitirem a utilização de termos semelhantes nas strings e (3) pelo número de artigos obtidos devido à abrangência das bases de dados utilizadas. Além desses critérios, a relevância dessas bibliotecas para a área de pesquisa deste trabalho foi crucial para as escolhas.

- **Idiomas dos artigos:** os idiomas escolhidos foram o Inglês e Português. Inglês, por ser o idioma mais utilizado pelos periódicos e conferências internacionais na área dessa pesquisa. Português, por ser o idioma nativo do pesquisador.
- **Termos utilizados na pesquisa:** para a definição de palavras-chaves utilizadas na string de busca foi aplicado o critério PICOC (*Population, Intervention, Comparison, Outcome e Context*), conforme sugerido por Kitchenham e Charters (2007). Neste MSL, a comparação e contexto (*Comparison e Context*) não foram aplicados devido ao fato desta pesquisa não ter como objetivo a comparação de tecnologias, mas sim a caracterização delas. Sendo assim, o PICOC foi definido da seguinte forma:

- *Population* (P): Interfaces Naturais de Usuário;
- *Intervention* (I): tecnologias que avaliam a Usabilidade e/ou a UX utilizadas no desenvolvimento de software que implementam NUIs;
- *Comparison* (C): não se aplica;
- *Outcome* (O): avaliação de Usabilidade/UX do software que utiliza uma NUI;
- *Context* (C): não se aplica.

Na Tabela 3.4 são exibidos os termos e a *string* de busca utilizada na pesquisa. Os termos são divididos em três partes: a primeira apresenta a população, ou seja, os termos relacionados à NUI; a segunda representa a intervenção, o que se pretende descobrir; e a terceira apresenta os resultados, ou seja, o que se deseja avaliar ou melhorar.

Tabela 3.4: Termos e *string* de busca utilizados no MSL.

<i>String</i> de busca	
População	("natural user interface*"OR "natural interface*"OR "natural user interaction*"OR "natural user communication*"OR "natural communication")
Intervenção	("tool"OR "framework"OR "technique"OR "method"OR "model"OR "process"OR "guideline"OR "pattern"OR "metric"OR "approach"OR "inspection"OR "principle"OR "aspect"OR "requirement"OR "heuristic"OR "methodology"OR "mechanism")
Resultados	("usability evaluation"OR "usability assessment"OR "usability improvement"OR "ux evaluation"OR "ux assessment"OR "ux improvement"OR "user experience evaluation"OR "user experience assessment"OR "user experience improvement")

3.1.4 Critérios de Seleção de Artigos

Todos os artigos retornados por este MSL foram avaliados independentemente por dois pesquisadores com o objetivo de incluir ou excluir no processo de seleção. De acordo com Kitchenham e Charters (2007), a utilização de dois ou mais pesquisadores revisando em conjunto os artigos é importante para que se mantenha uma consistência e diminua o viés da revisão. Antes de processo de seleção, os pesquisadores definiram em conjunto os critérios de inclusão e exclusão que são listados a seguir.

- **Critérios para a inclusão de artigos:**

- CI1: Publicações que apresentam tecnologias que avaliam a Usabilidade e/ou a UX no processo de desenvolvimento de software que implementa NUI;
- CI2. Publicações onde são descritos estudos experimentais das tecnologias para avaliar a Usabilidade e/ou a UX no processo de desenvolvimento de software que implementa NUI;
- CI3. Publicações que discutam aspectos relacionados a avaliação da Usabilidade e/ou da UX no processo de desenvolvimento de software que implementa NUI.

- **Critérios para a exclusão de artigos:**

- CE1. Não foram selecionadas publicações que não atendam aos critérios de inclusão;
- CE2. Não foram selecionadas publicações que não tenham disponibilidade de conteúdo para leitura e análise dos dados (especialmente em casos onde os artigos são pagos ou não são disponibilizados pelas máquinas de buscas);
- CE3. Não foram selecionadas publicações que possuem idioma diferente de Inglês e Português;
- CE4. Não foram selecionadas publicações que fazem parte da literatura cinzenta (publicações ou arquivos que não foram revisados por pares de revisores), como por exemplo relatórios técnicos, livros e trabalhos em progresso;
- CE5. Não foram selecionadas publicações que já foram incluídas em outra máquina de busca definida neste MSL (duplicadas).

3.1.5 Processo de Seleção de Artigos

No processo de seleção preliminar, chamado de primeiro filtro (1º Filtro), os dois pesquisadores avaliaram o título e o resumo dos artigos retornados, e os classificaram em incluídos ou excluídos de acordo com os critérios listados acima. Caso os pesquisadores não conseguissem alcançar uma conclusão sobre o artigo com base apenas em seu título e resumo, o

artigo era automaticamente incluído para ser avaliado na próxima etapa. Caso o artigo fosse excluído, era necessário apresentar uma justificativa plausível para a exclusão do mesmo.

No processo de seleção final, denominado segundo filtro (2º Filtro), os pesquisadores realizaram a leitura completa dos artigos que foram aprovados no 1º filtro. Os pesquisadores utilizaram os mesmos critérios de seleção para incluir ou excluir os artigos. Se o artigo fosse selecionado no 2º Filtro, seguia para a fase de extração de dados. Da mesma forma, caso o artigo não passasse para a etapa de extração, era necessário apresentar uma justificativa para sua exclusão.

3.1.6 Definição de estratégia de extração de dados

A estratégia de extração de dados designada neste mapeamento foi baseada em fornecer respostas para cada subquestão de pesquisa definida acima. A estratégia certifica a aplicação dos mesmos critérios de extração de dados para todos os artigos selecionados pelo MSL, facilitando a classificação.

Nesta etapa do MSL, todos os artigos que foram incluídos no 2º filtro tiveram seus dados extraídos. A extração dos dados de cada artigo foi revisada pelo autor e sua orientadora e disponibilizada para consulta no link: <https://bit.ly/3iD0WXf>. Para cada artigo aprovado foi criada uma tabela de extração de dados contendo as respostas de acordo com as subquestões. Além dos tópicos dessas questões, foram extraídos também os seguintes aspectos: dispositivo de captação de NUI utilizado pelo software, se avaliação foi realizada com foco em algum grupo específico de pessoa e descrição dos resultados da avaliação, como é mostrado na Tabela 3.5.

Tabela 3.5: Modelo de tabela para extração de dados.

Referência do artigo (Nome, ano). Título.	
Informações sobre a tecnologia de avaliação de Usabilidade e/ou UX	
SQ1. A tecnologia avaliava qual tipo de interação de NUI?	A tecnologia podia avaliar as seguintes interações, segundo (Vetere et al., 2014): (i) Multitouch : utiliza dois ou mais dedos no touchscreen para realizar alguma ação, por exemplo a “pinça” do zoom; (ii) Gesto : utilização de movimentos gestuais para realizar alguma ação, por exemplo o movimento de “bater” para clicar sobre alguma opção; (iii) Voz : utilização da voz para realizar alguma ação, por exemplo um comando de salvar um novo lembrete; (iv) Olhar : utilização de rastreamento ocular para realizar alguma ação, por exemplo afixação do olho em alguma letra de um teclado virtual
Qual dispositivo de captação foi utilizado no software que implementou a NUI?	<i>Kinect, Leap Motion, Microfone, EyeTracker</i> , entre outros.

SQ2. Qual o critério de qualidade da tecnologia de avaliação?	A tecnologia podia apresentar os seguintes critérios de qualidade: (i) Usabilidade : a tecnologia tem como objetivo avaliar a Usabilidade do software; (ii) UX : a tecnologia tem como objetivo avaliar a experiência de quem utiliza o software; (iii) Ambas : a tecnologia tem como objetivo avaliar tanto a Usabilidade quanto a UX do software.
SQ2.1 Quais os tipos de tecnologia de avaliação de Usabilidade foram usados?	As tecnologias de avaliação de Usabilidade foram classificadas de acordo com Ivory e Hearst (2001): (i) Teste : utilização de um avaliador observando os usuários utilizarem o sistema para encontrar possíveis falhas na Usabilidade; (ii) Inspeção : utilização de um especialista para identificar possíveis falhas na Usabilidade; (iii) Investigação : utilização de um método de coleta de dados dos participantes; (iv) Modelagem analítica : utilização de modelos da engenharia para prever a Usabilidade; (v) Simulação : utilização de algoritmos para simular a interação do usuário.
SQ2.2 Quais os tipos de tecnologia de avaliação de UX foram usados?	As tecnologias foram classificadas de acordo com Roto et al. (2009): (i) Estudo de laboratório : disponibilização do sistema em um contexto simulado para perceber as experiências dos usuários; (ii) Estudo de caso : disponibilização do sistema em um contexto real para perceber as experiências dos usuários; (iii) Survey : coleta de informações online dos usuários para compreender e perceber suas experiências; (iv) Especialista : disponibilização do sistema a um especialista para detectar possíveis falhas de UX.
SQ2.3 Quais aspectos de Usabilidade e/ou UX a tecnologia avalia?	As respostas obtidas por essa SQ são subjetivas e variam de estudo para estudo. No entanto, o objetivo principal foi constatar quais foram os aspectos avaliados pela tecnologia, como por exemplo eficácia, satisfação de usuário, fadiga, entre outros.
SQ3. A tecnologia é específica para software que implementa NUI ou é para software em geral?	A tecnologia podia ser classificada em um dos grupos: (i) Específica : a tecnologia de avaliação da Usabilidade e/ou UX é específica para software que implementa NUI; (ii) Genérica : a tecnologia de avaliação da Usabilidade e/ou UX não é específica para software que implementa NUI.
SQ4. A avaliação utiliza uma tecnologia de avaliação já existente?	A tecnologia podia ser classificada em uma das seguintes respostas: (i) Sim , a avaliação é realizada por meio de uma tecnologia de avaliação já existente. Qual?; (ii) Não , a avaliação não é realizada por meio de uma tecnologia de avaliação já existente.
Descrição da tecnologia	Descrever a tecnologia de avaliação

Informações sobre a avaliação empírica	
SQ5. O software que implementa a NUI foi avaliado experimentalmente?	A resposta podia se enquadrar em uma das seguintes: (i) Sim , o software foi avaliado experimentalmente; (ii) Não , o software não foi avaliado experimentalmente.
A avaliação foi realizada com algum grupo específico de pessoas?	A resposta poderia se enquadrar em dois grupos: (i) Sim . Qual grupo?; (ii) Não .
SQ5.1 Como e por qual tecnologia de avaliação foi realizada a avaliação empírica?	O objetivo da resposta desta SQ é detalhar como foi realizado o experimento do estudo, suas etapas, métodos e qual a tecnologia usada para avaliar a Usabilidade e/ou a UX.
Resultado da avaliação	Descrever quais foram os resultados obtidos com a avaliação empírica.
SQ5.2 A avaliação empírica foi feita na academia, indústria ou laboratório?	O ambiente de avaliação pode ser classificado em uma das seguintes respostas: (i) Industrial : a avaliação do sistema foi realizada em um contexto industrial e com profissionais; (ii) Acadêmico : a avaliação do sistema foi realizada em um contexto acadêmico e com alunos; (iii) Laboratório : a avaliação do sistema foi realizada em um contexto laboratorial; (iv) Misto : a avaliação do sistema foi realizada tanto no contexto industrial, quanto no acadêmico, ou laboratorial.
SQ5.3 A avaliação empírica foi analisada de maneira quantitativa ou qualitativa?	A avaliação podia ser classificada em uma das seguintes respostas: (i) Qualitativa : análise da avaliação feita de forma qualitativa; (ii) Quantitativa : análise da avaliação feita forma quantitativa; (iii) Mista : análise da avaliação feita de forma tanto qualitativa, quanto quantitativa.

3.1.7 Artigos selecionados após a condução do MSL

Conforme exibido na Tabela 3.6 abaixo, foram retornados 246 artigos após a aplicação da string de busca nas máquinas de busca selecionadas. Um total de 126 artigos foram selecionados após a aplicação do 1º filtro, com base nos critérios de inclusão e exclusão mostrados anteriormente. Um total de 56 artigos foram selecionados após a aplicação do 2º filtro (ver Apêndice A). Alguns artigos apareceram mais de uma vez em bibliotecas diferentes. Nestes casos, eles foram considerados apenas na primeira biblioteca de acordo com a ordem de busca realizada: Scopus, IEEEExplore, Science Direct, Engineering Village e ACM.

Tabela 3.6: Total de artigos retornados e selecionados no 1º e 2º filtro.

Fonte	Total de artigos retornados	Artigos selecionados no 1º filtro	Artigos aceitos no 2º filtro
Scopus	138	76	38
IEEEExplore	67	39	15
Science Direct	31	11	3
Engineering Village	7	0	0
ACM Digital Library	3	0	0
Total	246	126	56

Conforme mostrado na tabela acima, foram extraídos os dados dos 56 artigos aprovados no 2º filtro, os quais pautaram os resultados deste MSL.

3.2 RESULTADOS OBTIDOS

3.2.1 Visão geral dos resultados obtidos

Os resultados deste MSL, que são baseados na resposta de cada subquestão de pesquisa, são apresentados na Tabela 3.7. O resumo mais detalhado das respostas estão no Apêndice E. No geral, foram retornadas 110 tecnologias de avaliação utilizadas nos estudos, que serviram de análise para as subquestões SQ2, SQ3 e SQ4. A SQ1 apresenta 119 tecnologias e isso se deve ao fato de que a mesma tecnologia pode ter sido utilizadas diversas vezes no mesmo artigo para avaliar diferentes tipos de NUI, aumentando o número na contagem de 110 para 119 tecnologias.

Já nas subquestões SQ5, SQ5.2 e SQ5.3 foram levadas em consideração o número de avaliações empíricas realizadas. A SQ5 mostra uma análise booleana em relação aos 56 artigos que apresentam avaliações empíricas. No entanto, o número total de avaliações retornadas foi de 64 nas subquestões SQ5.2 e SQ5.3, e isso se deve ao fato de que no mesmo artigo, mais de uma avaliação empírica ocorreu, aumentando o número total da análise.

Tabela 3.7: Resultados do MSL para cada uma das subquestões.

Subquestões de Pesquisa	Possíveis respostas	Resultados	
		Tecnologias	Porcentagem (%)
SQ1. A tecnologia avalia qual tipo de NUI?	Multitoque	14	11,76
	Gesto	80	67,23
	Voz	17	14,29
	Olhar	8	6,72
SQ2. Qual o critério de qualidade da tecnologia de avaliação?	Usabilidade	94	85,45
	UX	14	12,73
	Ambas	2	1,82

Tabela 3.7: Resultados do MSL para cada uma das subquestões.

Subquestões de Pesquisa	Possíveis respostas	Resultados	
		Tecnologias	Porcentagem (%)
SQ2.1 Quais os tipos de tecnologia de avaliação de Usabilidade foram usados?	Teste	40	41,67
	Inspeção	2	2,08
	Investigação	54	56,25
	Modelagem analítica	0	0
	Simulação	0	0
SQ2.2 Quais os tipos de tecnologia de avaliação de UX foram usados?	Estudo de laboratório	16	100
	Estudo de caso	0	0
	Survey	0	0
	Especialista	0	0
SQ3. A tecnologia é específica para software que implementa NUI?	Sim	30	27,27
	Não	80	72,73
SQ4. A avaliação utiliza uma tecnologia de avaliação já existente? Se sim, qual?	Sim	82	74,55
	Não	28	25,45
	Total de tecnologias	110	100
		Avaliações	Porcentagem (%)
SQ5. O software que implementa a NUI foi avaliado experimentalmente?	Sim	55	98,21
	Não	1	1,79
SQ5.2 A avaliação foi feita na academia, indústria ou laboratório?	Indústria	0	0
	Laboratório	61	95,31
	Academia	1	1,56
	Mista	2	3,13
SQ5.3 A avaliação empírica foi analisada de maneira quantitativa ou qualitativa?	Quantitativo	36	56,25
	Qualitativo	9	14,06
	Ambas	19	29,69
	Total de avaliações	64	100

Sendo assim, este MSL identificou 110 tecnologias que foram utilizadas para avaliar a Usabilidade e/ou a UX em softwares que implementaram uma NUI. É importante ressaltar que esse número envolve duplicatas, ou seja, uma tecnologia pode ter sido usada mais de uma vez na contagem. Retirando as duplicatas e separando as tecnologias sem repetição, este MSL

encontrou 30 tecnologias de avaliação diferentes. Estas tecnologias estão descritas no Apêndice B. É possível observar que as subquestões SQ2.3 (Quais aspectos de Usabilidade e/ou UX a tecnologia avalia?) e SQ5.1 (Como e por qual tecnologia de avaliação foi realizada a avaliação empírica?) não foram apresentadas na Tabela 3.7. A subquestão SQ2.3 não foi apresentada devido à grande quantidade de respostas variadas que ela identificou. A subquestão SQ5.1, além do motivo anterior, também não foi apresentada na Tabela 3.7 devido ao fato dela ser uma questão subjetiva e variar de artigo para artigo.

3.2.2 Ano de publicação

Os artigos selecionados foram publicados entre 2011 e 2019. Conforme mostrado na Figura 3.1, houve um aumento do número de publicações no ano de 2013. Percebe-se também, de acordo com os artigos coletados nesse MSL, que o número de publicações diminuiu em 2018. O ano de 2019 não pôde ser analisado integralmente pois este MSL considerou artigos publicados até março de 2019. Isso pode justificar o baixo número de publicações desse ano. O ano de 2017 é o ano com maior quantidade de artigos identificados (13 artigos), seguido de 2016 e 2014 (10 artigos cada).

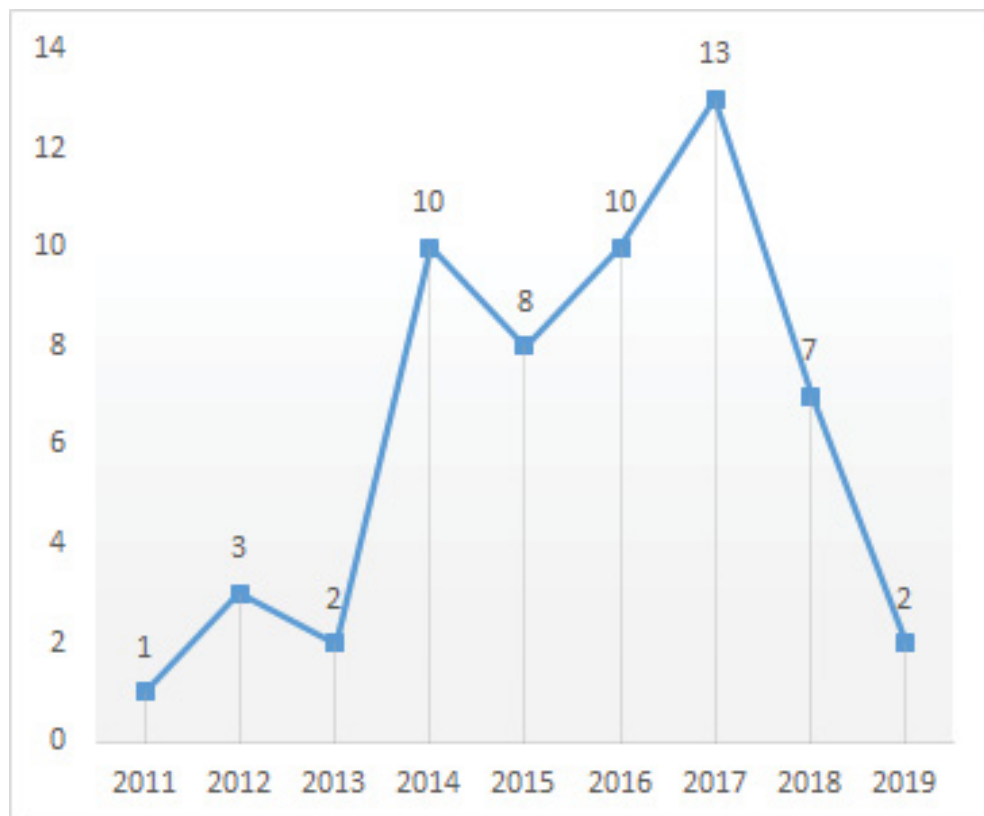


Figura 3.1: Ano de publicação dos artigos selecionados.

3.2.3 Locais de publicação

Neste MSL, foram considerados apenas locais de publicação avaliados por pares (incluindo periódicos, conferências e congressos). A Figura 3.2 fornece uma visão geral da distribuição de artigos por conferências. As conferências com maior número de publicação são *International Conference on Human-Computer Interaction – Interacción* (AIPO), *International Conference on Advanced Intelligent Mechatronics* (AIM), *International Conference on Intelligent User Interfaces* (IUI) e *International Conference on Systems, Man, and Cybernetics* (SMC), com dois artigos cada.

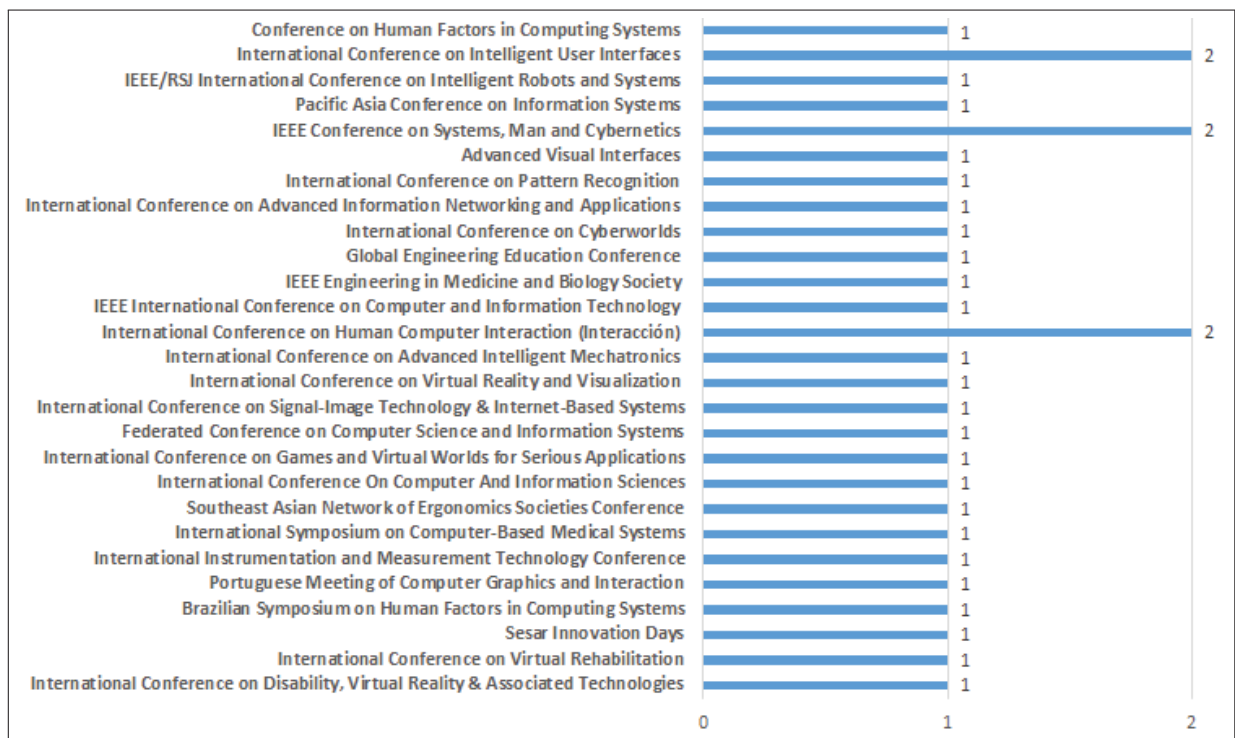


Figura 3.2: Distribuição de artigos por conferência.

A Figura 3.3 apresenta uma visão geral dos artigos de periódicos identificados. Os principais periódicos são *International Journal of Human-Computer Studies* (JHCS), *Universal Access in the Information Society* (UAIS) e *Interacting with Computers* (IWC), com dois artigos cada.

Os congressos onde houveram publicações são *Digital Heritage International Congress* (DHIC) e *World Congress on Health and Biomedical Informatics* (MEDINFO), com um artigo cada.

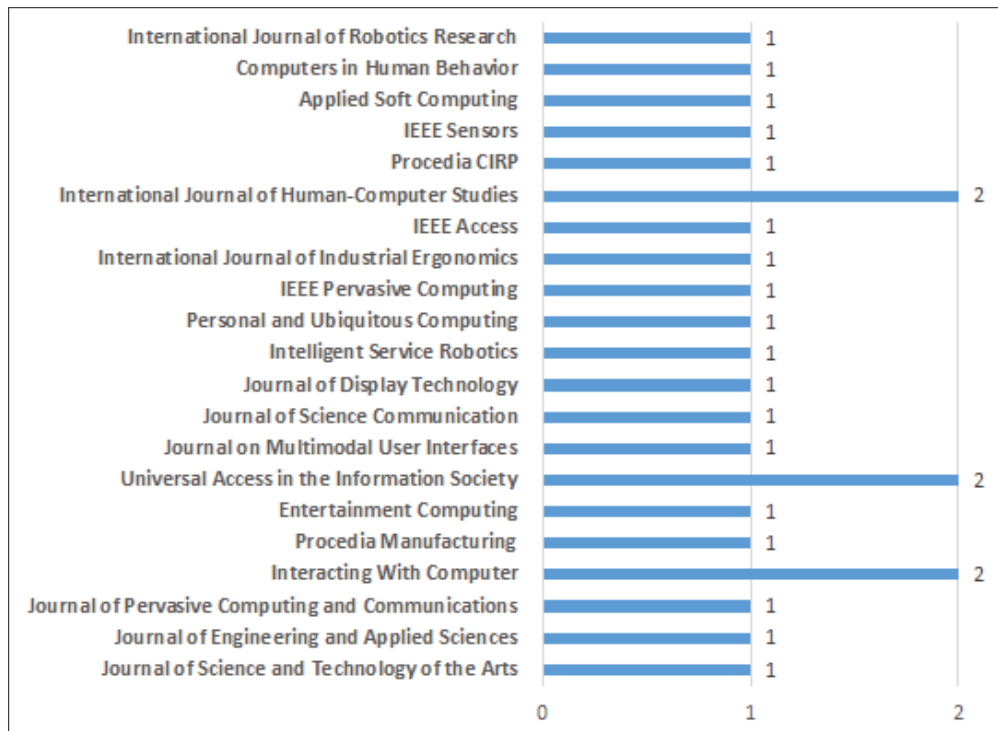


Figura 3.3: Distribuição de artigos por periódico.

3.2.4 Tipo de interação da NUI avaliada (SQ1)

Os resultados dessa subquestão revelaram que 67,23% (N = 80) das tecnologias utilizadas nos artigos avaliaram a Usabilidade e/ou a UX de softwares que utilizaram interfaces baseadas em gesto (como um website que permite rolar o *scroll* de sua página com um gesto vertical de mão). Por exemplo, Falcão *et al.* (2015) apresentam uma avaliação de Usabilidade de um experimento que continha tarefas a serem feitas no *Photoshop* utilizando gestos de mãos. Neste caso, a tecnologia de avaliação utilizada foi um questionário que combinava as Heurísticas de Nielsen (1993) com as de Jordan (1998). Estas heurísticas focam no projeto de uma interface com Usabilidade, onde as de Nielsen (1993) envolvem questões como controle e liberdade para o usuário, prevenção de erros e ajuda e documentação, enquanto as de Jordan (1998) focam em questões como coerência, clareza visual e explicitação.

Cerca de 14,29% (N = 17) das tecnologias retornadas avaliaram a Usabilidade e/ou a UX de softwares que utilizaram interfaces baseada em voz (como uma ferramenta de busca que realiza uma pesquisa de acordo com a palavra dita pelo usuário). Um exemplo de avaliação de Usabilidade desse tipo de interface foi realizado em Rocha *et al.* (2017), onde os autores executam um experimento no qual a tarefa principal foi realizar uma pesquisa por meio da voz. Neste estudo, uma das tecnologias de avaliação utilizada foi a observação direta.

Cerca de 11,76% (N = 14) das tecnologias retornadas foram utilizadas para avaliar a Usabilidade e/ou UX de softwares que utilizaram interfaces baseadas em multitoque (como um aplicativo de mapas que utiliza dois dedos para aumentar o *zoom* da navegação). Por exemplo, no estudo realizado por Uebbing-Rumke *et al.* (2014), foi avaliada a UX de um software controlador

de voo que era manuseado por meio do multitoque. Nesse caso, a tecnologia de avaliação utilizada foi o questionário *User Experience Questionnaire* (UEQ) (Laugwitz et al., 2008).

Por fim, das tecnologias retornadas, 6,72% (N = 8) avaliaram a Usabilidade e/ou a UX de software que implementaram a interface baseada em olhar (como um programa computacional que possui seu cursor controlado por meio do olhar). Por exemplo, Zhu *et al.* (2011) avaliaram a Usabilidade de um jogo virtual de pebolim, onde os usuários utilizavam o olhar para controlar os jogadores e chutar a bola para o gol. Nesse estudo, uma das tecnologias de avaliação utilizadas foi a gravação e análise do vídeo do experimento.

Os resultados dessa subquestão indicam que há necessidade de realizar mais estudos que utilizam a voz, olhar e multitoque como forma de interação. Além disso, é possível perceber que a grande maioria dos estudos que envolvem NUI utilizam os gestos para realizar a interação entre usuário e sistema.

3.2.5 Dispositivo para a captação de interações naturais

A Figura 3.4 apresenta uma visão geral dos dispositivos utilizados para captar as interações naturais dos usuários.

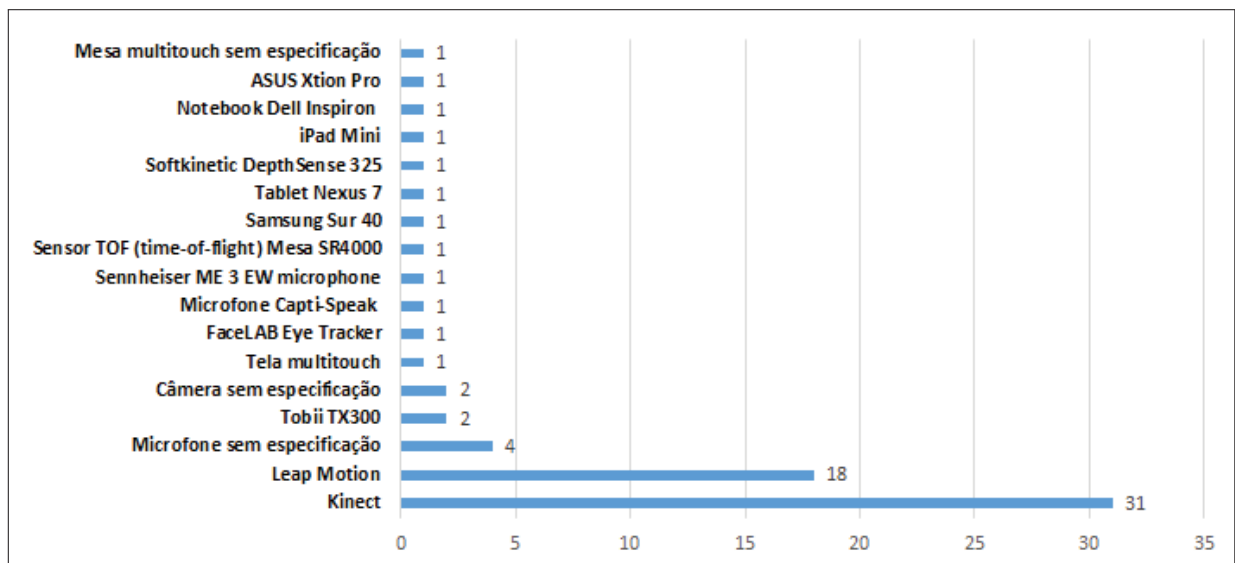


Figura 3.4: Dispositivos de captação utilizados nos estudos.

Os resultados dessa subquestão mostraram que 31 estudos utilizaram o *Kinect* para captar os gestos dos usuários. O *Kinect* é um captador de movimentos desenvolvido pela *Microsoft* (2010), e utilizado em seus *videogames*. A característica principal do *Kinect* é a alta precisão para captar gestos do corpo inteiro do usuário, além da facilidade para desenvolver aplicativos que o utilize, devido ao *Software Development Kit* (SDK) do *Kinect* para *Windows*, desenvolvido pela *Microsoft* (2014). Um exemplo de uso desse dispositivo é apresentado em Kazuma *et al.* (2016), o qual os autores propuseram e avaliaram um novo modelo de apresentações de *slides* baseado em gestos captados do *Kinect*.

O *Leap Motion* é o segundo dispositivo mais utilizado (em 18 estudos). É um dispositivo de captação de gestos, no entanto, esse sensor capta apenas gestos dos dedos das mãos (Leap Motion, 2013). Um exemplo de uso desse dispositivo é mostrado em Vosinakis *et al.* (2016), onde os autores desenvolveram um aplicativo no qual os usuários, por meio da interação com *Leap Motion*, assumiam o papel de escultores e conseguiam progressivamente criar uma estátua virtual, selecionando e aplicando as ferramentas apropriadas.

Os resultados dessa questão mostram que diversos dispositivos têm sido utilizados para captar as interações naturais dos usuários. O alto uso do *Kinect* e *Leap Motion* é justificado porque o gesto é a forma de interação mais utilizada. A popularidade e qualidade desses dispositivos também justifica a escolha dos desenvolvedores quando vão programar suas aplicações. É possível observar também que, quando é utilizada a voz, olhar ou multitoque, não há algum dispositivo que se destaque na quantidade de uso ou característica positiva.

3.2.6 Critério de qualidade da tecnologia de avaliação (SQ2)

Os resultados dessa subquestão indicam que 85,45% (N = 94) das tecnologias de avaliação possuem como foco a Usabilidade. Um exemplo do uso é apresentado em Deng *et al.* (2017), onde os autores avaliaram a satisfação do usuário, um dos atributos da Usabilidade, ao utilizar um software que tinha como objetivo usar gestos e olhar para mover cubos à posições-alvo. Neste caso, a tecnologia de avaliação utilizada foi o *System Usability Scale* (SUS) (Brooke, 1996).

Cerca de 12,73% (N = 14) das tecnologias de avaliação retornadas no MSL possuíam foco na UX. O trabalho de d'Ornellas *et al.* (2015) apresenta uma avaliação de aspectos como imersão, tensão, desafio e competência dos usuários ao utilizarem um “jogo sério” por meio de gestos para a reabilitação de derrame cerebral. A tecnologia de avaliação de UX utilizada foi o módulo principal do *Game Experience Questionnaire* (GEQ) (Ijsselstein *et al.*, 2013).

Por fim, cerca de 1,82% (N = 2) das tecnologias avaliaram tanto a Usabilidade quanto a UX. Um exemplo de uso é apresentado em Economou *et al.* (2017), no qual os autores avaliaram (por meio da mesma tecnologia de avaliação) a satisfação (Usabilidade) e o engajamento dos usuários (UX) ao utilizarem um “hangout virtual” com avatares captados pelo *Kinect*. Neste caso, foi utilizada a gravação do experimento e análise posterior dos vídeos gravados para obter as informações sobre estes critérios de qualidade.

Os resultados da subquestão SQ2 mostram que há uma carência de tecnologias que avaliam em conjunto a Usabilidade e a UX especificamente de softwares que implementam alguma NUI. Sendo assim, os aspectos pragmáticos de Usabilidade são avaliados em uma ocasião, e os aspectos hedônicos de UX em outra, faltando tecnologias que unam essas duas vertentes.

Dentro da subquestão SQ2, há uma subquestão chamada SQ2.1 (Tipo de avaliação de Usabilidade). Os resultados de SQ2.1 mostram que as tecnologias de investigação são as mais usadas, cerca de 56,25% (N = 54). Esse tipo de tecnologia utiliza artefatos para coleta dados dos participantes do experimento, como por exemplo o questionário de satisfação apresentado em Rybarczyk *et al.* (2018), onde os usuários poderiam respondê-lo após utilizar um sistema de

tele-reabilitação. Cerca de 41,67% (N = 40) das tecnologias de avaliação são do tipo teste de Usabilidade, onde é o pesquisador observa a interação dos usuários no experimento ou avaliando os resultados alcançados por eles. Por exemplo, em Tang *et al.* (2018) o pesquisador utilizou uma ficha de observação para registrar qualquer movimento peculiar durante a realização das tarefas. Ainda, cerca de 2,08% (N = 2) das tecnologias são de inspeção, onde é necessário um avaliador especialista para encontrar problemas de Usabilidade. Por exemplo, em Guimarães *et al.* (2012) os pesquisadores desenvolveram uma ficha de inspeção com diversos aspectos a fim de avaliar o sistema. Por fim, não houveram tecnologias de modelagem analítica e simulação.

Os resultados da subquestão SQ2.1 mostram que as tecnologias de avaliação mais utilizadas são as de investigação, as quais coletam dados dos usuários do experimento. Nenhuma tecnologia de modelagem ou simulação foi encontrada neste MSL. Isso pode ter acontecido pelo fato desses tipos de tecnologias dependerem de outras tecnologias, modelos e suporte ferramental.

Ainda sobre a subquestão SQ2, há uma subquestão SQ2.2 (Tipo de avaliação de UX). De acordo com os resultados da SQ2.2, todas as tecnologias de avaliações foram de estudo de laboratório. Isso mostra que essa tecnologia é a mais simples de ser aplicada no contexto de NUIs, visto que o estudo de caso necessita de ambiente real de aplicação, o survey depende uma avaliação online e o estudo com especialista necessita de um expert na área, tornando esses métodos mais complicados de serem executados.

Por fim, há a subquestão SQ2.3 (Aspectos avaliados de Usabilidade e/ou UX pela tecnologia de avaliação). A Figura 3.5 apresenta uma visão geral dos aspectos retornados neste MSL.

3.2.7 Tecnologias de avaliação específicas ou não (SQ3)

Os resultados dessa subquestão mostraram que 72,73% (N = 80) das tecnologias de avaliação tem como objetivo avaliar pra software em geral, ou seja, não são específicas para avaliar software que implementa NUI. Como por exemplo, em Caggianese *et al.* (2014), onde os autores utilizam um questionário de Usabilidade que pode ser usado em diferentes contextos, o *System Usability Scale* (SUS) (Brooke, 1996), para avaliar um software que manipulava esculturas virtuais por meio de gestos com as mãos.

Cerca de 27,27% (N = 30) das tecnologias foram criadas para avaliar especificamente os sistemas que implementaram NUIs. Em Su *et al.* (2014), por exemplo, os autores desenvolveram um questionário para avaliar a Usabilidade de um sistema baseado em gesto para reabilitação domiciliar. O questionário continha dez questões para ser respondidas usando uma escala *Likert* de 4 pontos, no entanto, era específico para o trabalho que foi desenvolvido, utilizando até o nome do software que foi avaliado.

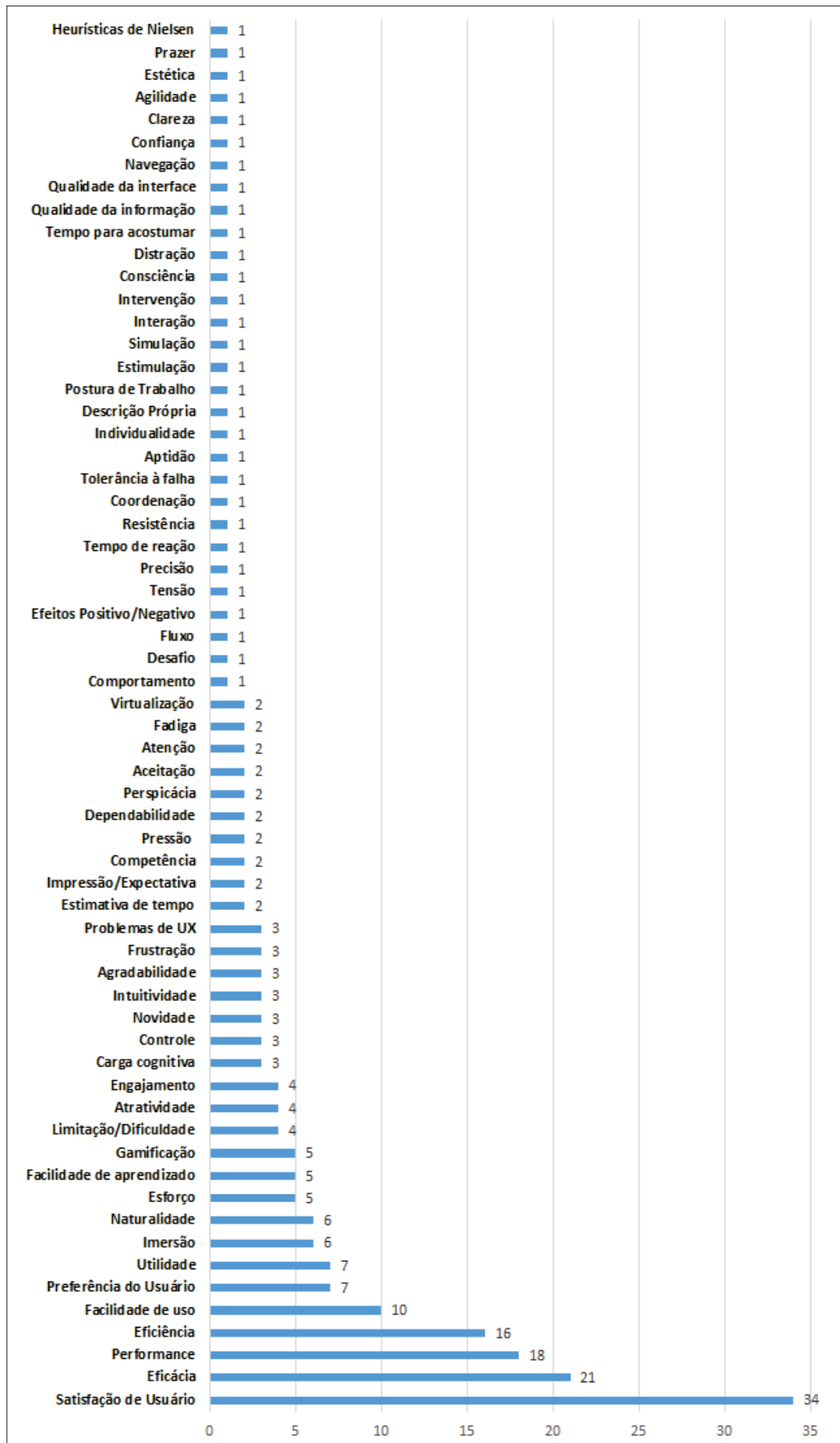


Figura 3.5: Aspectos de Usabilidade e UX avaliados pelas tecnologias.

Os resultados indicam que há uma carência de tecnologias específicas para avaliar Usabilidade e/ou UX de softwares que implementam algum tipo de NUI. Isso porque quando não é usada uma tecnologia que seja replicável, são desenvolvidas tecnologias utilizadas apenas em um trabalho, fazendo com que a avaliação de Usabilidade/UX desse tipo de software não seja específica para o seu contexto e especificidade. Isso faz com que sejam utilizadas tecnologias genéricas para o contexto de NUI, o que pode fazer com que a avaliação não seja boa o suficiente visto que as tecnologias genéricas não englobam aspectos específicos deste tipo de interação que precisam ser avaliados.

3.2.8 Uso de tecnologias baseadas em outras já existentes ou não (SQ4)

Os resultados dessa subquestão mostram que 74,55% (N = 82) das tecnologias de avaliação utilizadas já tinham sido criadas anteriormente ao estudo executado. Por exemplo, em Uebbing-Rumke *et al.* (2014) os autores utilizam o *User Experience Questionnaire* (UEQ) (Laugwitz *et al.*, 2008), para realizar a avaliação de UX.

Cerca de 25,45% (N = 28) das tecnologias de avaliação não se baseiam em outra existente, ou seja, os autores criaram uma nova tecnologia de avaliação para realizada a avaliação da NUI. É o caso de Macaranas *et al.* (2015), onde os autores desenvolveram uma entrevista com questões direcionadas para avaliar a Usabilidade do software do estudo.

A Figura 3.6 mostra a combinação dos resultados das subquestões SQ3 (Tecnologias de avaliação específicas ou não) e SQ4 (Uso de tecnologias de avaliação já existentes ou não). Esses resultados indicam que houveram 30 tecnologias de avaliação para NUI e 80 tecnologias de avaliação genéricas. Isso mostra que o uso das tecnologias genéricas, no contexto de NUI, é muito superior em relação às tecnologias específicas. Ainda, das 30 tecnologias específicas, apenas 6 se basearam em alguma existente, ou seja, as outras 24 tecnologias foram desenvolvidas apenas para avaliar a NUI em questão, sem fornecer um padrão comum de avaliação e replicabilidade.

3.2.9 Avaliação empírica (SQ5)

Os resultados dessa subquestão mostram que apenas um artigo não possuiu avaliação empírica. Neste caso, Sun e Cheng (2014) mostraram o software que desenvolveram e a tecnologia de avaliação, no entanto, não realizaram um estudo empírico.

Dentro da SQ5, há as subquestões SQ5.1 (Qual tecnologia de avaliação utilizada no estudo empírico?), SQ5.2 (Qual o ambiente da avaliação?) e SQ5.3 (Qual o tipo de análise?). Os resultados gerais da SQ5.1 são mostrados na Figura 3.7 abaixo.

Os resultados dessa subquestão mostram que o SUS e a análise das pontuações do estudo foram as tecnologias de avaliação mais utilizadas (19 vezes cada). Um exemplo do uso dessas tecnologias pode ser visto em Li *et al.* (2016) e Eckert *et al.* (2017), respectivamente. Em Li *et al.* (2016), os pesquisadores utilizam o SUS para avaliar a Usabilidade da interação a distância em telas de exibição. Com esse questionário é possível avaliar a satisfação do usuário

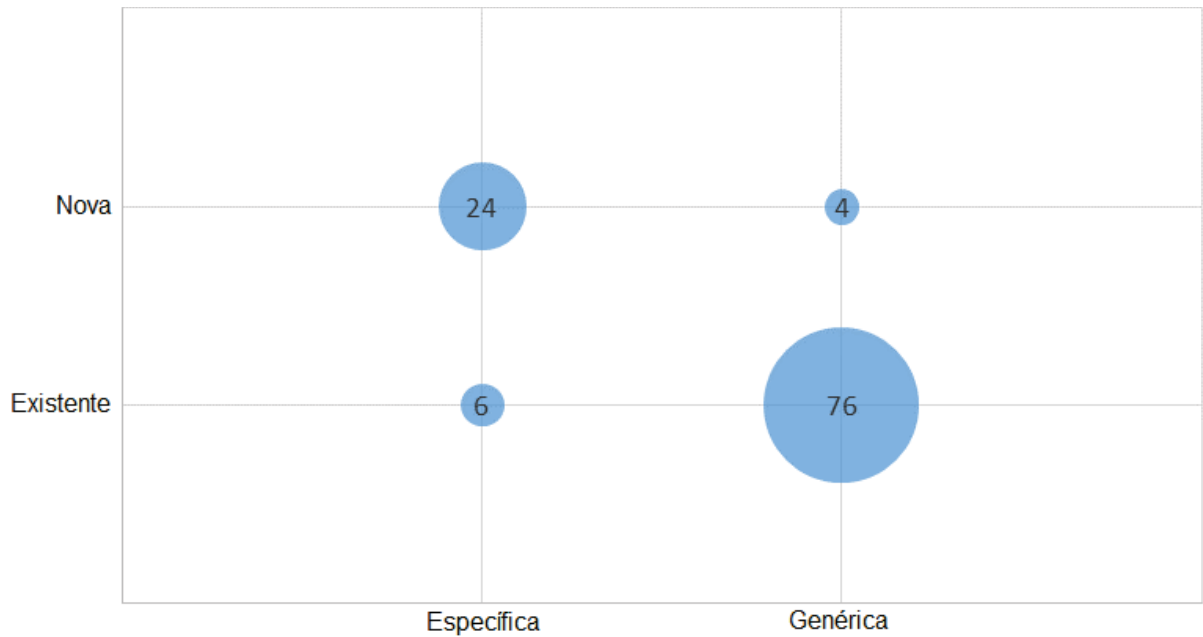


Figura 3.6: Resultados combinados das subquestões SQ3 e SQ4.

por meio de 10 questões que podem ser respondidas com a escala *Likert*. Em relação ao estudo descrito em Eckert *et al.* (2017), os autores analisaram as pontuações do estudo para avaliar a Usabilidade. Neste caso, a eficiência e a eficácia dos participantes foram obtidas de acordo com análise dos resultados de performance e pontuação.

Os resultados da subquestão SQ5.2 mostraram que 95,31% (N = 61) dos estudos empíricos foram realizados em ambiente laboratorial, como em Su *et al.* (2014), o qual os autores simularam um contexto para testar um sistema de reabilitação em casa por meio do *Kinect*. Cerca de 3,13% (N = 2) foram realizados em ambiente misto, como é o caso de Shishido *et al.* (2015), que mesclaram ambiente laboratorial com acadêmico. Neste caso, os autores simularam um contexto de utilização do sistema em um ambiente acadêmico, feito com alunos. Apenas Zhu *et al.* (2011) realizaram o estudo exclusivamente na academia, o qual os autores realizaram o experimento em um contexto real com estudantes universitários. Nenhum dos estudos identificados neste MSL foi feito na indústria. Com isso, é possível concluir que criar um ambiente laboratorial para aplicar o estudo tornou-se mais viável no contexto de NUI. Isso se deve ao fato de que em um ambiente simulado é possível desenvolver e conduzir mais facilmente experimentos controlados com NUI. Além disso, é mais interessante avaliar primeiramente um produto em um contexto simulado do que em um contexto real, para que seja possível encontrar melhorias e atualizações no produto. É importante também mencionar a ausência de estudos realizados na indústria, visto que é necessário realizar parcerias com empresas de desenvolvimento para aplicar as tecnologias, o que dificulta a realização do experimento. No entanto, os estudos realizados na indústria são importantes para verificar a aplicação da tecnologia em um contexto real de uso.

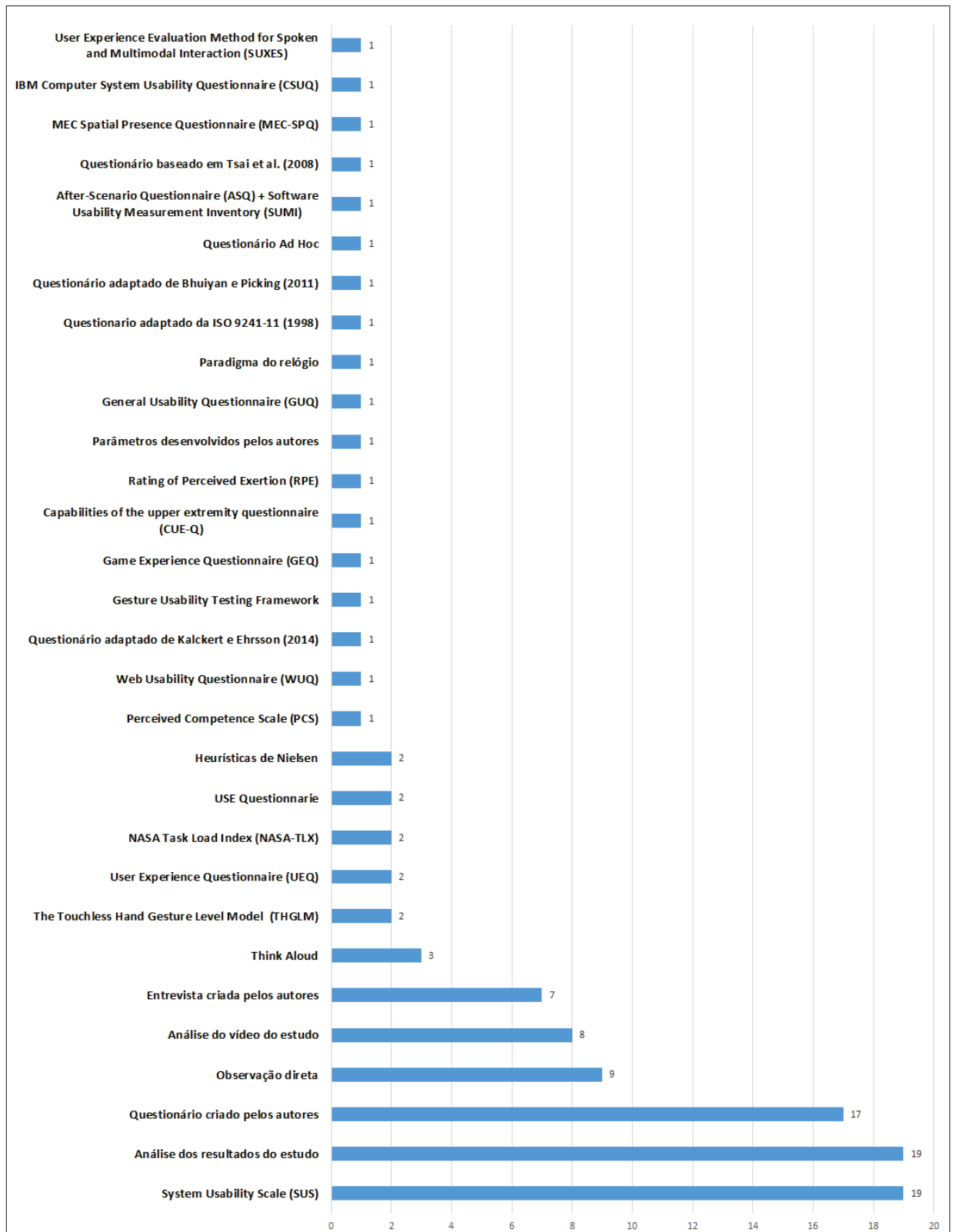


Figura 3.7: Tecnologias de avaliação utilizadas nos estudos.

Os resultados da subquestão SQ5.3 revelaram que 56,25% (N = 36) das análises foram realizadas quantitativamente. Por exemplo, em Lee *et al.* (2017) os autores (por meio de gráficos) mostraram os resultados obtidos com as respostas do questionário aplicado para avaliar a NUI.

Além disso, cerca de 29,69% (N = 19) das análises foram feitas de forma mista (quantitativamente e qualitativamente). Por exemplo, em Chatzidaki e Xenos (2017) os autores utilizaram o tempo de realização das tarefas para avaliar a eficiência (quantitativo) e as respostas da entrevista para avaliar as opiniões dos participantes (qualitativo). Por fim, somente 14,06% (N = 9) foram realizadas de forma qualitativa, valorizando a subjetividade das respostas, como em Economou *et al.* (2017), onde os autores, por meio da análise da gravação do experimento, conseguiram extrair informações subjetivas sobre os aspectos definidos por eles, como o foco de atenção do usuário. Com esses resultados é possível constatar que há uma escassez de tecnologias que avaliem qualitativamente os resultados das avaliações realizadas, focando na maioria das vezes apenas nos números. É necessário que haja um equilíbrio na realização dos dois tipos de análises para obter uma análise da tecnologia mais consistente.

3.3 SÍNTESE DO CAPÍTULO

Este capítulo abordou os resultados obtidos de um estudo secundário (MSL), e mostrou as evidências encontradas nas bibliotecas digitais sobre as tecnologias que têm sido usadas para avaliar a Usabilidade e/ou UX em softwares que implementam as NUIs. A partir de um total inicial de 246 artigos, 56 foram selecionados e extraídos neste mapeamento, após aplicação do 1º e 2º filtro.

Os resultados deste MSL identificaram 30 tecnologias de avaliação diferentes e suas características que têm sido utilizadas para avaliar os softwares que implementam as NUIs, com o objetivo de verificar se o sistema fornece uma boa Usabilidade e uma boa experiência aos seus usuários.

Algumas lacunas foram identificadas neste MSL. Uma delas diz respeito à necessidade de se criar uma tecnologia de avaliação de Usabilidade e/ou de UX que avalie especificamente softwares que implementam NUIs e que seja replicável, podendo ser utilizada em diversos contextos que essas aplicações estão inseridas, como por exemplo agentes conversacionais, assistentes virtuais, aplicativos de mapa ou aplicativos em geral que utilizam comando de voz. Com isso, a avaliação da Usabilidade e/ou de UX desses softwares será direcionada a eles, fornecendo resultados específicos que podem auxiliar na melhoria destas aplicações.

Ainda, entre os tipos de interação das NUIs mostrados no mapeamento, foi notado que algumas interações têm sido pouco avaliadas em âmbitos gerais, como é o caso da voz. Portanto, foi verificada a possibilidade de focar esta pesquisa nesse tipo de interação, visto que ela vem ganhando espaço entre os aplicativos e sistemas nos últimos anos, aumentando a repercussão entre os usuários. Além disso, tem-se notado a necessidade de propor novas maneiras de avaliação para pesquisadores e desenvolvedores que produzem esse tipo de interação de NUI.

4 PROPOSTA DA PRIMEIRA VERSÃO DA U2X ECS

Este capítulo apresenta a proposta da primeira versão de uma tecnologia que tem como objetivo avaliar a Usabilidade e UX de sistemas que utilizam a voz como forma de interação.

4.1 INTRODUÇÃO

O desenvolvimento desta proposta foi pautado de acordo com os resultados obtidos no Mapeamento Sistemático da Literatura descrito no Capítulo 3. Como pôde ser visto, muitas tecnologias diferentes são utilizadas para avaliar sistemas que utilizam um tipo de interação para NUI. No entanto, poucas tecnologias de avaliação são especificamente para esse tipo de interface. Em grande parte das vezes, os autores utilizam tecnologias que podem ser usadas para avaliar qualquer tipo de interface, como Caggianese *et al.* (2014), Uebbing-Rumke *et al.* (2014) e Su *et al.* (2014).

Em relação à Usabilidade e UX, foi possível perceber que há uma grande preocupação dos pesquisadores em avaliar essas perspectivas em seus trabalhos. No entanto, essa avaliação é realizada de forma separada utilizando tecnologias diferentes para cada um desses conceitos, o que exige muita dedicação por parte do pesquisador ao decidir avaliar os dois conceitos em um único trabalho. Por exemplo, em Deng *et al.* (2017) é avaliado apenas a Usabilidade e em d'Ornellas *et al.* (2015) apenas a UX. É possível pensar no mesmo sentido quando o assunto é o tipo de análise. Os pesquisadores se importam com avaliações que forneçam resultados quantitativos e qualitativos, mas geralmente obtém resultados em apenas uma destas vertentes. Por exemplo, Lee *et al.* (2017) realizam a análise da avaliação apenas de forma quantitativa, e Economou *et al.* (2017) que realizam sua análise somente de forma qualitativa.

No que se refere aos tipos de interação das Interfaces Naturais de Usuário foi observado que algumas delas foram pouco avaliadas quando comparadas às outras, como a interação baseada em voz. Além disso, não foi identificada uma tecnologia específica para este tipo de interação. Portanto, foi vista a possibilidade de focar esta pesquisa nesse tipo de interação, com o objetivo de promover e melhorar a comunicação do usuário com o sistema utilizando a voz.

As lacunas observadas e comentadas acima mostram a carência de uma tecnologia de avaliação de Usabilidade em conjunto com UX, que avalie tanto quantitativamente quanto qualitativamente a interação baseada em voz. De acordo com essa necessidade, a proposta inicial desta pesquisa é a definição de uma tecnologia de avaliação chamada *Usability and User Experience Evaluation of Conversational Systems* (U2X ECS). A U2X ECS é uma tecnologia baseada em questionário. Este tipo de tecnologia de coleta foi escolhido pois é consolidado na literatura e por meio dele pode-se extrair tanto dados quantitativos quanto qualitativos.

4.2 ANÁLISE DAS TECNOLOGIAS DE AVALIAÇÃO IDENTIFICADAS NO 1º MSL

Uma análise das tecnologias identificadas no primeiro MSL que avaliaram a voz foi realizada para verificar quais dessas tecnologias poderiam contribuir para a construção dessa proposta. Para isso, uma leitura minuciosa das tecnologias foi realizada a fim de verificar seus itens de avaliação e sua aplicabilidade no contexto da voz. A Tabela 4.1 a seguir apresenta uma análise dessas tecnologias e seus pontos que podem ser considerados ou não por esta proposta.

Tabela 4.1: Tecnologias que retornaram no MSL que avaliaram uma interação baseada em voz

Tecnologia	Descrição
<i>System Usability Scale</i> – SUS (Brooke, 1996)	O SUS é uma tecnologia para medir a Usabilidade de produtos/serviços. É um questionário de 10 itens que pode ser respondido por meio de uma escala <i>Likert</i> de 5 pontos, onde 1 representa “discordo fortemente” e 5 representa “concordo fortemente”. Mesmo não sendo específica para a interação baseada em voz, é uma tecnologia que possui itens e aspectos de Usabilidade que podem ser utilizados no contexto da voz, como a eficiência, eficácia e satisfação do usuário. Sendo assim, o SUS foi considerado útil para a formulação desta proposta.
<i>Software Usability Measurement Inventory</i> – SUMI (Kirakowski e Corbett, 1993)	O SUMI é um questionário de 50 itens que se destina a ser utilizado com uma amostra de usuários que participaram de algum experimento com um software. As respostas para os itens podem variar em “concordo”, “indeciso” e “discordo”. O conceito de Usabilidade, tal como avaliado pelo SUMI, baseia-se na definição da norma ISO 9241. Por ser um questionário com um conjunto de itens grande, mesmo não sendo focado para interação baseada em voz ele foi considerado útil para a formulação desta proposta.

Tabela 4.1: Tecnologias que retornaram no MSL que avaliaram uma interação baseada em voz

Tecnologia	Descrição
Questionário autoral (Di Nuovo et al., 2018)	O questionário foi desenvolvido para avaliar a Usabilidade de constructos que estavam sendo testados em um experimento. É um questionário que continha 7 questões que podiam ser respondidas por meio de uma escala <i>Likert</i> de 5 pontos, onde 1 representava “discordo” e 5 representava “concordo”. Por ter questões criadas especificamente para avaliação da voz como interação, suas questões foram consideradas úteis para esta proposta.
Entrevista autoral (Ashok et al., 2017)	Foi realizada uma entrevista com 3 questões abertas com os participantes para gerar comentários e sugestões sobre um experimento. No entanto, os autores não expuseram as questões utilizadas e, com isso, as questões da entrevista não puderam ser consideradas na formulação desta proposta.
Questionário autoral (Lee et al., 2017)	O questionário foi utilizado pelos autores para avaliar aspectos de Usabilidade. As respostas poderiam variar em uma escala <i>Likert</i> de 7 pontos, onde 1 era um oposto negativo sobre o aspecto e 7 representava o oposto positivo. Mesmo sendo um questionário de Usabilidade que pode ser utilizado em diferentes contextos além da interação por voz, os autores detalham a boa estrutura que o questionário segue. Sendo assim, ele foi considerado útil na formulação da proposta.
Questionário autoral (Gustavsson et al., 2017)	Os autores desenvolveram e aplicaram o questionário para verificar a opinião sobre a Usabilidade da interação baseada em voz. O questionário continha 11 questões sobre aspectos de interesse e utilidade, e podia ser respondido por meio de uma escala <i>Likert</i> de 5 pontos, onde 1 representava “discordo fortemente” e 5 “concordo fortemente”. O questionário continha uma estrutura relevante acerca da avaliação da interação baseada em voz e, por isso, foi considerado útil na formulação desta proposta.

Tabela 4.1: Tecnologias que retornaram no MSL que avaliaram uma interação baseada em voz

Tecnologia	Descrição
Questionário autoral (Profanter et al., 2015)	O questionário desenvolvido pelos autores buscou avaliar a expectativa e impressão que os participantes tiveram ao utilizar algumas modalidades de interação, entre elas a voz. Para isso, os participantes podiam responder como “muito mais fácil que o esperado”, “como esperado” e “muito mais complicado do que o esperado”. Como o questionário possuiu certa especificidade ao avaliar a voz, ele foi considerado útil nesta proposta.
Questionário autoral (Profanter et al., 2015)	Os autores desenvolveram o questionário para comparar a Usabilidade de diferentes modalidades de interação utilizadas no estudo. Por meio da preferência do participante, eles classificaram as modalidades de acordo com as tarefas realizadas. Por ser uma tecnologia de avaliação comparativa, ou seja, comparava por exemplo a interação por voz com a interação baseada por toque, ela não foi considerada útil para a construção desta proposta.
Entrevista autoral (Chatzidaki e Xenos, 2017)	Os autores realizaram uma entrevista com os participantes após a execução do experimento, com o objetivo de verificar a UX. No entanto, a estrutura e questões da entrevista não foram detalhadas e, por isso, não foi considerada na elaboração desta proposta.
Observação direta (Rocha et al., 2017; Ashok et al., 2017)	A observação direta consiste basicamente de um ou mais pesquisadores utilizando ou não uma estrutura de observação durante a execução do experimento com o objetivo de registrar possíveis desvios de padrão. O pesquisador responsável por observar deve anotar percepções como por exemplo reações dos participantes e opiniões dadas durante o experimento para possível análise posterior. Esse tipo de avaliação difere do estilo questionário que foi seguido por esta proposta, portanto não foi considerada na elaboração.

Tabela 4.1: Tecnologias que retornaram no MSL que avaliaram uma interação baseada em voz

Tecnologia	Descrição
Análise do vídeo do experimento (Rocha et al., 2017)	A análise do vídeo do experimento é uma forma de avaliação que se baseia na gravação visual de todo o experimento. Os pesquisadores gravam com uma câmera toda execução e, após o experimento, visualizam o vídeo diversas vezes para captar possíveis aspectos que considerem importante. Como essa tecnologia difere do estilo questionário adotado para este trabalho, ela não foi considerada na proposição.
Análise das pontuações do experimento (Gustavsson et al., 2017; Chatzidaki e Xenos, 2017)	A análise das pontuações do experimento consiste nos pesquisadores obterem conclusões a partir de números alcançados no experimento. Por exemplo, realizar uma análise de eficácia a partir do score alcançado pelos participantes no experimento ou ainda, analisar a eficiência por meio do tempo levado para executar uma tarefa. Essa tecnologia de avaliação varia do estilo questionário proposto por esse trabalho e não foi considerada na elaboração.

As tecnologias mostradas acima foram todas retornadas pelo MSL descrito no Capítulo 3. Como pode ser visto, algumas tecnologias não foram utilizadas devido à especificidade de cada uma, por exemplo a não especificação das questões e a diferença nos tipos de tecnologia. Paralelamente, outras tecnologias influenciaram a formulação desta proposta. Desse modo, afim de fundamentar ainda mais a proposta da tecnologia U2XECS, outras duas tecnologias de avaliação de Usabilidade que não foram retornadas no MSL foram consideradas. Essas tecnologias não foram retornadas no MSL devido ao fato de que elas foram criadas antes de popularização do termo Interface Natural de Usuário. São elas:

- SASSI (Hone e Graham, 2000): o questionário de Avaliação Subjetiva de Interfaces de Sistema de Fala (SASSI – *Subjective Assessment of Speech System Interfaces*) concentra-se na qualidade da entrada de fala. É um conjunto de 34 itens respondidos por escala *Likert* que avaliam seis conjuntos de aspectos: precisão da resposta do sistema, satisfação do usuário, demanda cognitiva, aborrecimento, "*habitability*" e velocidade. Diferentemente dos outros questionários, o SASSI também avalia a "*habitability*", que "se refere à medida que o usuário sabe o que fazer e sabe o que o sistema está fazendo" (Hone e Graham, 2000).
- Checklist autoral de Farinazzo *et al.* (2010): um *checklist* de Usabilidade baseado nas heurísticas de Nielsen e adaptadas para o contexto de interação baseada em voz. O

checklist foi desenvolvido para ser usado por especialistas e contém um conjunto de 14 categorias que foram consideradas nessa proposta.

Como pode ser observado, nenhuma tecnologia de avaliação de UX foi utilizada para avaliar a voz e retornada no MSL. Além disso, mesmo procurando abertamente na literatura por diferentes modos (busca no Google, *snowballing*, Research Gate), não foi encontrada nenhuma tecnologia de UX para avaliação específica da interação por voz. Sendo assim, foram selecionadas as métricas de UX propostas por Bargas-Ávila e Hornbaek (2011) e comentadas por outros autores para embasar a vertente de avaliação de UX fornecida pela proposta. A escolha dessas métricas foi por conta do seu uso para a formulação de tecnologia a realização de estudos na literatura. Além disso, para complementar nas escolhas das palavras utilizadas nas questões, foi verificado também as métricas de qualidade hedônica avaliadas pelo questionário *AttrakDiff* (Hassenzahl et al., 2003). A escolha das métricas criadas por esses autores se dá pelo fato de que esses trabalhos possuem grande relevância e importância na área, com utilizações em outras pesquisas. As métricas são:

- *UX Genérica*: não há foco específico em nenhum aspecto da UX. Refere-se a uma impressão ou sentimento do uso geral ou experiência com o sistema. Um exemplo de avaliação dessa métrica para Polkosky (2005) seria utilizando palavras relacionadas a “bom sentimento” ou “boa experiência” do usuário ao utilizar sistema;
- *Afeto/Emoção*: afeto refere-se a estados psicológicos, incluindo emoções, sentimentos, impressões e humores. Emoção surge como resultado da interação com um produto ou sistema. Bargas-Ávila e Hornbaek (2011) afirmam que o afeto e a emoção são tratados, na maioria das vezes, como sinônimos na pesquisa da UX, como uma explicação para colocá-los na mesma dimensão. Esses tipos de estados já foram abordados antigamente por Mehrabian e Russel (1974) ao os relacionarem com os sentimentos das pessoas como felicidade, prazer, satisfação, relaxamento, estimulação, excitação e admiração;
- *Prazer/Diversão*: esta dimensão refere-se a interações lúdicas caracterizadas por percepções de prazer e envolvimento. Quão divertido usar um sistema é o foco dessa métrica. Para Lavie e Tractinsky (2004), pode-se avaliar essa métrica por meio dos sentimentos dos usuários tais como espontaneidade, imaginação, criatividade, felicidade, originalidade e inovação;
- *Estética/Atração*: esta dimensão refere-se à estética clássica, enfatizando o design limpo e ordenado e a estética expressiva associada às qualidades de criatividade e novidade. Características físicas e sensoriais de um produto ou uma interação que resulta em atratividade. Hassenzahl (2008) aborda essas métricas por meio de atributos de produto ou de interação como agradável, convidativo, atraente, simpático, motivador e desejável.

- *Engajamento/Fluxo*: Engajamento refere-se a “uma categoria de UX caracterizada por atributos de desafio, afeto positivo, durabilidade, apelo estético e sensorial, atenção, *feedback*, variedade/novidade, interatividade e controle do usuário percebido” (O’Brien e Toms, 2008). Fluxo se refere a “um estado em que as pessoas estão tão envolvidas em uma atividade que nada mais parece importar; a experiência em si é tão agradável que as pessoas farão isso mesmo com grande custo, pelo simples prazer de fazê-lo” (Jackson e Marsh, 1996). Para estes autores, fluxo pode ser traduzido pelos sentimentos das pessoas como controle, foco de atenção, curiosidade e interesse intrínseco;
- *Motivação*: esta dimensão refere-se a “fatores internos que impulsionam a ação e a fatores externos que podem agir como incentivos à ação”. Para O’Brien e Toms (2008) são basicamente os atributos que encorajam ou desencorajam os usuários;
- *Encantamento*: esta dimensão refere-se a um estado em que um sentimento de desorientação coexiste com níveis elevados de percepção e atenção. Para McCarthy *et al.* (2006), pode-se avaliar o encantamento a partir da imersão obtida pelo usuário ao utilizar determinado produto ou interação;

4.3 PRIMEIRA VERSÃO DA U2XECS

A U2XECS é uma tecnologia que avalia a Usabilidade e UX de sistemas conversacionais que utilizam a voz para realizar interações. A U2XECS é composta por um questionário, que tem por objetivo extrair dados quantitativos e qualitativos de variados aspectos relacionados a essas duas vertentes da IHC, a Usabilidade e a UX, e pode ser usado tanto por avaliadores/desenvolvedores no momento de avaliação de seus sistemas, quanto para pesquisadores que queiram investigar os aspectos que esta tecnologia abrange.

A utilização da U2XECS é simples e direta. Basicamente o avaliador deve solicitar que usuários respondam ao questionário logo após a utilização do sistema de interação por voz que se quer avaliar. As questões quantitativas são respondidas de acordo com uma escala *Likert* de 5 pontos, onde 1 representa “eu não concordo com essa afirmação” e 5 representa “eu concordo com essa afirmação”. Foi escolhido este tipo de escala pois ela fornece diversas possibilidades de escolha para o participante e, além disso, uma resposta central caso ele não consiga se decidir. Já as respostas qualitativas são abertas e dependem da motivação do participante para escrevê-la, podendo variar de uma simples palavra até um texto exaustivo.

As questões de Usabilidade foram baseadas nos questionários encontrados na literatura por meio do MSL e nas demais fontes identificadas e adaptadas para o contexto da pesquisa. A U2XECS segue os aspectos definidos na ISO 9241-11 (2011) para avaliar a Usabilidade, que são a eficácia, a eficiência e a satisfação do usuário. Já para a avaliação da UX, as questões foram todas criadas com base nas métricas de UX propostas de Bargas-Ávila e Hornbaek (2011).

A Tabela 4.2 abaixo apresenta as questões abordadas pela primeira versão da U2XECS e as referências utilizadas para a sua criação. As questões foram abordadas de uma forma que os usuários as compreendam corretamente e saibam exatamente o que estão avaliando.

Tabela 4.2: Aspectos, questões e referências utilizadas na primeira versão da U2XECS.

Aspecto	Questão	Referências
Satisfação de usuário	1 - É fácil utilizar a voz para realizar as tarefas nesse sistema.	SUS (Brooke, 1996); SASSI (Hone e Graham, 2000); (Lee et al., 2017; Di Nuovo et al., 2018; Gustavsson et al., 2017).
	2 - Preciso aprender muito antes de realizar essas tarefas com a voz.	SUS (Brooke, 1996); SASSI (Hone e Graham, 2000).
	3 - É satisfatório usar a voz para realizar essas tarefas.	SUMI (Kirakowski e Corbett, 1993); (Lee et al., 2017; Gustavsson et al., 2017).
	4 - A interação por voz me deixou desconfiante para usar esse sistema.	SUS (Brooke, 1996); SASSI (Hone e Graham, 2000); SUMI (Kirakowski e Corbett, 1993); (Farinazzo et al., 2010).
Eficiência	5 - A interação por voz causa efeito imediato no sistema.	SUMI (Kirakowski e Corbett, 1993)
	6 - Eu usaria a voz com frequência para realizar essas tarefas.	SASSI (Hone e Graham, 2000); SUMI (Kirakowski e Corbett, 1993); (Farinazzo et al., 2010)
Eficácia	7 - O sistema falhou ao reconhecer o que eu falei.	SUS (Brooke, 1996); SASSI (Hone e Graham, 2000)
	8 - A voz foi útil para realizar essas tarefas.	SASSI (Hone e Graham, 2000); (Lee et al., 2017)
UX Genérica	9 - Realizar essas tarefas por meio da voz foi uma experiência boa.	(Bargas-Ávila e Hornbæk, 2011; Polkosky, 2005)
Afeto/Emoção	10 - Me senti animado ao utilizar a voz para realizar essas tarefas.	(Bargas-Ávila e Hornbæk, 2011; Mehrabian e Russel, 1974)
	11 - Foi muito prazeroso utilizar a voz para realizar essas tarefas.	(Bargas-Ávila e Hornbæk, 2011; Lavie e Tractinsky, 2004)

Tabela 4.2: Aspectos, questões e referências utilizadas na primeira versão da U2XECS.

Aspecto	Questão	Referências
Prazer / Diversão	12 - Foi muito maçante utilizar a voz para realizar essas tarefas.	(Bargas-Ávila e Hornbæk, 2011; Lavie e Tractinsky, 2004); Atrakdiff (Hassenzahl et al., 2003).
Estética / Atração	13 - Utilizar a voz fez com que eu me dedicasse mais para realizar as tarefas no sistema.	(Bargas-Ávila e Hornbæk, 2011; Hassenzahl et al., 2008); Atrakdiff (Hassenzahl et al., 2003).
	14 - Utilizar a voz para realizar as tarefas me deixou mais atraído e com vontade de usar o sistema.	(Bargas-Ávila e Hornbæk, 2011; Hassenzahl et al., 2008)
Engajamento / Fluxo	15 - O sistema possui um design inovador que facilita a realização das tarefas por meio da voz.	(Bargas-Ávila e Hornbæk, 2011; O'Brien e Toms, 2008)
	16 - Eu não consegui manter o foco nas tarefas ao utilizar a interação por voz para realizá-las.	(Bargas-Ávila e Hornbæk, 2011; Jackson e Marsh, 1996)
Motivação	17 - Me senti motivado ao utilizar a voz para realizar essas tarefas.	(Bargas-Ávila e Hornbæk, 2011; O'Brien e Toms, 2008)
Encantamento	18 - As ações que realizei no sistema com a voz fez com que eu ficasse totalmente cativado nas tarefas.	(Bargas-Ávila e Hornbæk, 2011; McCarthy et al., 2006); Atrakdiff (Hassenzahl et al., 2003).
Qualitativas	19 - No geral, utilizar a voz para realizar as tarefas foi... porque...	(Profanter et al., 2015)
	20 - No geral, um ponto positivo ao utilizar a interação por voz para realizar essas tarefas é... porque...	
	21 - No geral, um ponto negativo ao utilizar a interação por voz para realizar essas tarefas é... porque...	

4.4 TEMPLATE DO QUESTIONÁRIO DA PRIMEIRA VERSÃO DA U2XECS

QUESTIONÁRIO DA U2XECS
Nome:
<p>– Por favor, preencha o questionário abaixo para avaliar a sua experiência na interação com um software de interação por voz. Sua resposta é importante e nos ajudará a propor melhorias por meio desta avaliação.</p> <p>– Nas questões de múltipla escolha, assinale apenas uma resposta por questão. Nas questões dissertativas, fique à vontade para escrever o quanto desejar. Se necessário, utilize o verso dessas folhas.</p> <p>– Não há respostas certas ou erradas. O importante para nós é a sua opinião.</p> <p>– Responda as questões de múltipla escolha utilizando a seguinte escala:</p> <p>1 = discordo totalmente;</p> <p>2 = discordo parcialmente;</p> <p>3 = não discordo nem concordo;</p> <p>4 = concordo parcialmente;</p> <p>5 = concordo totalmente.</p>

	1	2	3	4	5
1 - É fácil utilizar a voz para realizar as tarefas nesse sistema.					
2 - Preciso aprender muito antes de realizar essas tarefas com a voz.					
3 - É satisfatório usar a voz para realizar essas tarefas.					
4 - A interação por voz me deixou confiante para usar esse sistema.					
5 - A interação por voz causa efeito imediato no sistema.					
6 - Eu usaria a voz com frequência para realizar essas tarefas.					
7 - O sistema falhou ao reconhecer o que eu falei.					
8 - A voz foi útil para realizar essas tarefas.					
9 - Realizar essas tarefas por meio da voz foi uma experiência boa.					
10 - Me senti animado ao utilizar a voz para realizar essas tarefas.					
11 - Foi muito prazeroso utilizar a voz para realizar essas tarefas.					
12 - Foi muito maçante utilizar a voz para realizar essas tarefas.					
13 - O sistema possui um design inovador que facilita a realização das tarefas por meio da voz.					
14 - Utilizar a voz para realizar as tarefas me deixou mais atraído e com vontade de usar o sistema.					

15 - Utilizar a voz fez com que eu me dedicasse mais para realizar as tarefas no sistema.					
16 - Eu não consegui manter o foco nas tarefas ao utilizar a interação por voz para realizá-las.					
17 - Me senti motivado ao utilizar a voz para realizar essas tarefas.					
18 - As ações que realizei no sistema com a voz fez com que eu ficasse totalmente cativado nas tarefas.					

19 - No geral, utilizar a voz para realizar as tarefas foi _____, porque:

20 - No geral, um ponto **positivo** ao utilizar a interação por voz para realizar essas tarefas é _____, porque:

21 - No geral, um ponto **negativo** ao utilizar a interação por voz para realizar essas tarefas é _____, porque:

4.5 ESTUDO EXPLORATÓRIO COM A PRIMEIRA VERSÃO DA U2XECS

Um estudo exploratório foi realizado com o objetivo de verificar a facilidade de uso, utilidade percebida e as intenções de uso da primeira versão da U2XECS. Buscou-se também, por meio deste estudo, captar possíveis orientações fornecidas pelos participantes a fim de refinar e evoluir a tecnologia.

4.5.1 Planejamento

O estudo foi realizado com dois potenciais usuários da U2XECS: um pesquisador e um desenvolvedor que trabalham no processo de desenvolvimento de software. A participante 01 é engenheira de software e possui experiência tanto na indústria como na academia. O participante

02 é aluno de pós-graduação, e trabalha com acessibilidade e jogos para pessoas com deficiência. Mesmo que os participantes não tenham desenvolvido ainda sistemas específicos que utilizam interação baseada em voz, eles os utilizam diariamente e são partes interessadas nesse tipo de sistema, pois trabalham ativamente no desenvolvimento dos mais variados tipos de software.

O Termo de Consentimento Livre e Esclarecido (TCLE) foi criado, distribuído aos participantes e coletado, a fim de assegurar a confidencialidade e os aspectos éticos que envolveram o estudo. Além disso, outros artefatos foram desenvolvidos para o estudo, como é o caso do questionário de caracterização e o treinamento para o uso da U2XECS. O primeiro artefato foi criado com o objetivo de captar as experiências prévias dos participantes para análise. O segundo foi utilizado para preparar os participantes para o estudo, para que soubessem o que seria feito e avaliado. Para avaliar a tecnologia U2XECS após o seu uso, foi utilizado o *Technology Acceptance Model* (TAM) (Venkatesh e Davis, 2000), que por meio de uma escala *Likert* de 7 pontos avalia aspectos de intenção de uso, utilidade percebida e facilidade de uso. Além dos aspectos avaliados pelo TAM, foram criadas três questões dissertativas para obter respostas mais detalhadas dos participantes sobre os pontos positivos e negativos, as dificuldades encontradas e as sugestões de melhoria. Ademais, foi criado também um roteiro que os participantes deveriam seguir, demonstrando as tarefas a serem realizadas no estudo e a ordem natural que ele devia seguir.

O aplicativo *Google Assistant*¹, em sua versão para dispositivos *mobile* foi utilizado neste estudo exploratório. A escolha desta aplicação se dá pelo crescimento dos assistentes virtuais desde 2017 e a forma que este tipo de aplicação está inserida no cotidiano da sociedade, auxiliando as pessoas nas mais variadas tarefas do dia a dia. Esse crescimento e a importância dessas aplicações na atualidade evidenciam a necessidade de realizar constantes avaliações e propor melhorias do seu uso que refletem na sociedade em geral.

4.5.2 Execução

O experimento foi realizado em dias diferentes com os participantes. Antes da realização das tarefas, o TCLE foi distribuído e assinado pelos participantes. Após a assinatura do TCLE, foi solicitado aos participantes que preenchessem o questionário de caracterização para análise posterior da experiência que eles possuíam em relação às aplicações que usam interação baseada em voz e à avaliação de Usabilidade e UX.

Em seguida, foi realizado o treinamento o qual foi apresentada a tecnologia U2XECS para introduzir o seu uso e aspectos. Além disso, foi explicado também o que seria feito durante o estudo e qual era o seu objetivo. Todas as dúvidas que surgiram relacionadas a qualquer aspecto foram imediatamente esclarecidas.

Foi distribuída então a folha de instrução, que continha os passos que os participantes do estudo exploratório deveriam seguir. Os passos eram:

¹<https://assistant.google.com/>

1. *Baixe o aplicativo Google Assistant pelo link ou pela sua loja de aplicativo;*
2. *Realize o login e finalize a configuração;*
3. *Anote aqui o horário inicial deste estudo: ____h____min;*
4. *Realize as seguintes tarefas no aplicativo Google Assistant:*
 - *Verifique o clima para o dia de hoje;*
 - *Coloque a música “Nothing Else Matters” para tocar no Youtube;*
 - *Feche o Youtube e pesquise a melhor rota para chegar no Shopping Jardim das Américas.*
5. *Após a finalização das tarefas, responda o questionário da U2XECS acerca da interação baseada em voz utilizada (Google Assistant);*
6. *Anote aqui o horário final deste estudo: ____h____min;*
7. *Responda ao Questionário Pós-uso TAM sobre o questionário da U2XECS;*

4.5.3 Análise

Venkatesh e Davis (2000) propuseram o TAM para verificar as respostas dos participantes sobre a aceitação ou rejeição das tecnologias. No caso deste estudo exploratório, os indicadores foram: (i) facilidade de uso percebida; (ii) utilidade percebida; e (iii) intenção de uso futura. Além disso, foram utilizadas três questões abertas para coletar dados qualitativos dos participantes.

Facilidade de uso percebida (FU) define o grau em que o participante acredita que usar a tecnologia seria livre de dificuldade, por meio das seguintes questões: (FU1) Minha interação com o questionário da U2XECS foi clara e compreensível; (FU2) Interagir com o questionário da U2XECS não exige muito do meu esforço mental; (FU3) Considero o questionário da U2XECS fácil de usar; e (FU4) Considero fácil de utilizar o questionário da U2XECS para fazer o que eu quero que ela faça, avaliar a Usabilidade e a UX do Sistema Conversacional.

Utilidade percebida (UP) define o grau em que o participante acredita que a tecnologia poderia melhorar seu desempenho, por meio das seguintes questões: (UP1) Usar o questionário da U2XECS melhorou o meu desempenho na avaliação da Usabilidade e a UX da interação por voz/análise das respostas dos participantes; (UP2) Usar o questionário da U2XECS permitiu aumentar a minha produtividade na avaliação da Usabilidade e a UX da interação por voz/análise das respostas dos participantes; (UP3) Usar o questionário da U2XECS aumentou a minha eficácia na avaliação da Usabilidade e UX da interação por voz/análise das respostas dos participantes; e (UP4) Considero o questionário da U2XECS útil para avaliação de Usabilidade e a UX da interação por voz/análise das respostas dos participantes.

Intenção de uso futuro (IU) define o grau em que o participante acredita que utilizaria a tecnologia em projetos futuros, por meio das seguintes questões: (IU1) Supondo que eu tenha

acesso ao questionário da U2XECS, eu pretendo usá-lo; e (IU2) Levando em conta que eu tenho acesso ao questionário da U2XECS, eu prevejo que irei usá-lo.

As Questões abertas (QA) foram utilizadas para coletar respostas dissertativas dos participantes sobre o uso do questionário: (QA1) Por favor, descreva os pontos positivos e negativos de fazer uma avaliação de Usabilidade e de UX com o questionário da U2XECS; (QA2) Você teve alguma dificuldade ao utilizar o questionário da U2XECS? Em caso afirmativo, quais dificuldades foram essas?; (QA3) Você poderia descrever sugestões de melhorias para o questionário da U2XECS ou sugestões de novos aspectos de avaliação que o questionário ainda não contempla?

Os participantes responderam o questionário em uma escala de sete pontos proposta por Venkatesh e Davis (2000). As possíveis respostas foram: concordo totalmente, concordo moderadamente, concordo parcialmente, neutro (não concordo nem discordo), discordo parcialmente, discordo moderadamente e discordo totalmente. Esta escala de respostas foi considerada adequada porque há um valor intermediário, ou seja, ela respeita a indecisão do participante fornecendo uma possibilidade de resposta neutra.

A Figura 4.1 apresenta a concepção dos participantes em relação ao questionário da U2XECS. O eixo vertical do gráfico apresenta afirmativas dos indicadores do TAM utilizados. Já o eixo horizontal representa o grau de aceitação dos participantes. Nas barras estão inseridos os códigos que simbolizam os participantes (P1 e P2) do estudo e suas respectivas avaliações. As respostas dos usuários mostram que o questionário possuiu boa aceitação. No entanto, é necessário atentar-se às intenções de uso do questionário, pois o P1 concordou parcialmente em utilizá-la posteriormente.

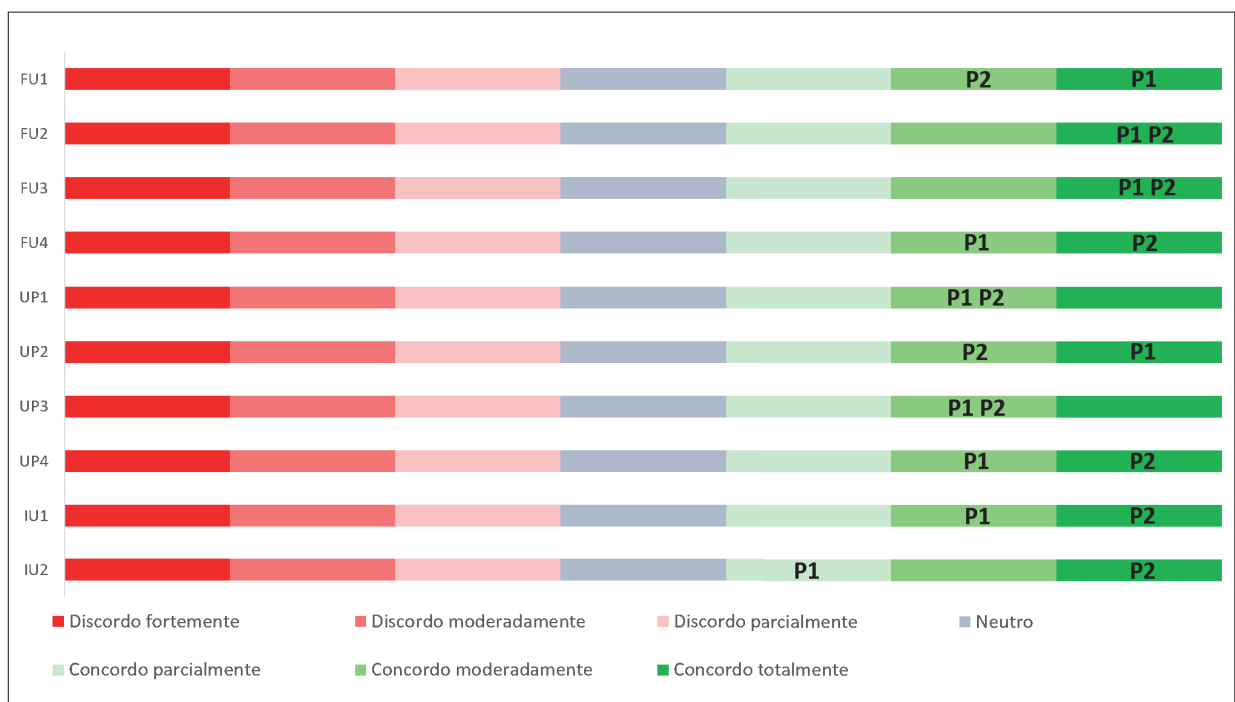


Figura 4.1: Concepção dos participantes em relação ao questionário da U2XECS.

Foi realizada também uma análise qualitativa das respostas das três questões dissertativas para verificar as opiniões dos participantes. O método utilizado foi o *Grounded Theory* (GT) (Strauss e Corbin, 2014). Foram utilizadas as duas primeiras fases do GT. Como o objetivo desta análise foi perceber a compreensão dos participantes sobre a U2XECS, não foi necessário executar a codificação seletiva (3ª fase), visto que apenas as duas primeiras fases já foram ideais para o objetivo proposto.

A análise qualitativa confirma a verificação positiva identificada na análise quantitativa por meio das categorias de qualidade. Para o questionário, resposta como “*o questionário é simples de usar*” (P1) e “*o questionário ressaltou questionamentos relevantes para avaliar esse tipo de tecnologia*” (P1) mostram o aceite dos participantes do estudo exploratório.

A análise de sugestões e alterações foram importantes para propor uma evolução do questionário. Essas análises mostraram opiniões dos usuários para melhorar o formato do questionário, como “*o questionário pode se aprofundar mais nos questionamentos da UX*” (P1) e “*o questionário pode ter questionamentos mais intrusivos a respeito do sentimento do usuário durante o uso*” (P1). Além disso, também identificaram melhorias para a escrita das questões, como “*deixar claro na questão 5 a que a palavra ‘efeito’ se refere*” (P2) e “*tirar a negação no meio da frase da questão 8*” (P2).

Tanto a análise quantitativa quanto a qualitativa tiveram respostas positivas. Isso nos leva a entender que os participantes gostaram da tecnologia. No entanto, há diversos pontos que precisam ser melhorados. As diversas respostas “concordo parcialmente” evidenciam a necessidade de propor melhorias e evoluções da tecnologia. Combinadas com as respostas qualitativas, foram propostas novas versões da U2XECS que abrangeram essas questões.

5 MAPEAMENTO SISTEMÁTICO DA LITERATURA SOBRE TECNOLOGIAS DE AVALIAÇÃO DE USABILIDADE E/OU UX PARA SCS

O primeiro MSL nos forneceu uma visão geral das tecnologias de avaliação de NUI, mas de uma forma abrangente. A partir da proposição da primeira versão da U2XECS e do estudo exploratório, foi vista a possibilidade de se realizar um segundo MSL, desta vez específico para SCS. Portanto, este capítulo apresenta um MSL das tecnologias de avaliação de Usabilidade e/ou UX usadas em SCS, abordando a seguinte questão de pesquisa: “*Quais tecnologias estão sendo usadas para avaliar a Usabilidade e UX de Sistemas Conversacionais?*”. Este capítulo também mostra como este MSL foi planejado e conduzido, além dos resultados e lacunas de pesquisa obtidos por meio dele.

5.1 PROTOCOLO DO MAPEAMENTO SISTEMÁTICO DA LITERATURA

5.1.1 Objetivo

O objetivo do MSL foi baseado no paradigma *Goal-Question-Metric* (GQM) (Basili e Rombach, 1988) e está descrito na Tabela 5.1.

Tabela 5.1: Objetivo do MSL.

Analisar	publicações contendo tecnologias que avaliam a Usabilidade e/ou UX de SCS
Com o propósito de	caracterizar
Em relação a	tecnologias de avaliação, SCS avaliados, e tipos de avaliações empíricas
Do ponto de vista dos	pesquisadores de IHC e ES
No contexto de	publicações disponíveis em Scopus, ACM, IEEE Xplore e Engineering Village

5.1.2 Questões de pesquisa

A principal questão de pesquisa deste MSL é “*Quais tecnologias estão sendo usadas para avaliar a Usabilidade e a UX de SCS?*”. De acordo com Santos et al. (2012), entende-se por “tecnologia” uma generalização de ferramentas, métricas, técnicas, metodologias, entre outras propostas nas áreas ES e IHC. Além da questão principal, foram definidas sub-questões (SQs) para responder a perguntas específicas sobre cada tecnologia de avaliação, os SCS avaliados em artigos, e a avaliação empírica realizada. Estas SQs são mostrados na Tabela 5.2.

Tabela 5.2: Subquestões de pesquisa do MSL.

Sub-questões de Pesquisa
SQ1. Qual é o critério de qualidade da tecnologia de avaliação utilizada?
SQ2. Quais aspectos da Usabilidade e/ou UX a tecnologia avalia?
SQ3. A tecnologia é específica para SCs ou para sistemas em geral?
SQ4. A tecnologia foi criada pelo estudo ou é baseada em uma tecnologia já existente?
SQ5. Como a tecnologia de avaliação coleta dados dos participantes?
SQ6. Quais são as características da tecnologia de avaliação?
SQ7. A tecnologia extrai dados quantitativos ou qualitativos?
SQ8. Qual é a função do SC?
SQ9. Qual categoria de aplicação do SC?
SQ10. O sistema foi criado para um grupo específico de pessoas? Quais grupos?
SQ11. A tecnologia de avaliação foi avaliada empiricamente?
SQ11.1. Em caso afirmativo, qual foi o tipo de experimento realizado para avaliar a tecnologia de avaliação?

5.1.3 Estratégia utilizada para pesquisa das publicações

A estratégia de busca fornece uma maior integridade à pesquisa (Kitchenham e Charters, 2007). Para isso, deve-se definir previamente a estratégia que será utilizada. A estratégia é descrita a seguir.

- **Escopo da pesquisa:** a pesquisa foi feita a partir de bibliotecas digitais através de seu mecanismo de busca avançada. A Tabela 5.3 mostra as bibliotecas utilizadas. Estas fontes de dados foram escolhidas porque elas: (i) fornecem um mecanismo de busca eficiente; (ii) permitem que a mesma *string* de busca seja usada; e (iii) retornam um número relevante de artigos. Além disso, a importância destas fontes de dados para nosso tópico de pesquisa foi um fator crucial na escolha.

Tabela 5.3: Bibliotecas de busca utilizadas.

Nome da fonte	Link
Scopus	http://www.scopus.com/home.url
ACM Digital Library	http://portal.acm.org/dl.cfm
IEEEExplore	https://ieeexplore.ieee.org/Xplore/home.jsp
Engineering Village	https://www.engineeringvillage.com/search/quick.url

- **Idiomas dos artigos:** os idiomas escolhidos foram o Inglês e Português. Inglês, por ser o idioma mais utilizado pelos periódicos e conferências internacionais na área dessa pesquisa. Português, por ser o idioma nativo do pesquisador.
- **Termos utilizados na pesquisa:** para definir as palavras-chave usadas na *string* de busca, foi aplicado o critério PICOC (*Population, Intervention, Comparison, Outcome* e *Context*) (Kitchenham e Charters, 2007). Em nosso MSL, as palavras-chave para Comparação e Contexto não foram utilizadas na string, pois o objetivo do MSL era caracterizar as tecnologias de avaliação, e não compará-las. Portanto, PICOC foi definido como segue:
 - (*P*)*opulation*: Sistemas Conversacionais/Interação Baseada em Voz;
 - (*I*)*Intervention*: tecnologias que avaliam a Usabilidade e/ou UX em SCs;
 - (*C*)*omparison*: não aplicável;
 - (*O*)*utcome*: avaliação de software através de tecnologias que verificam aspectos de Usabilidade e/ou UX;
 - (*C*)*ontext*: não aplicável.

Na Tabela 5.4 são mostrados os termos e a *string* de busca utilizada. Os termos estão agrupados em três divisões: a primeira parte representa a população: sistemas conversacionais e interação baseada em voz; a segunda parte representa a intervenção: o que se planeja encontrar, as tecnologias de avaliação; e a terceira parte representa os resultados: as avaliações de Usabilidade e/ou UX.

5.1.4 Critérios de seleção de artigos

Para classificar os artigos como incluídos ou excluídos, foram definidos critérios de inclusão e exclusão:

- Critérios para a inclusão de artigos:
 - CI1. Publicações contendo tecnologias que avaliam Usabilidade e/ou UX de SCs;

Tabela 5.4: Termos e *string* de busca utilizados no MSL.

<i>String</i> de busca	
População	("conversational system*"OR "voice interface*"OR "voice recognition"OR "voice-based interaction"OR "voice interaction"OR "speech interface*"OR "speech recognition"OR "dialogue interface*"OR "dialogue agent*"OR "speaker recognition*") AND
Intervenção	("tool" OR "framework" OR "technique" OR "method" OR "model" OR "process" OR "guideline" OR "pattern" OR "metric" OR "approach" OR "inspection" OR "principle" OR "aspect" OR "requirement" OR "heuristic" OR "methodology" OR "mechanism") AND
Resultados	("Usability evaluation"OR "Usability assessment"OR "ux evaluation"OR "ux assessment"OR "User eXperience evaluation"OR "User eXperience assessment"OR "user-centred evaluation")

- CI2. Publicações que descrevem estudos empíricos de tecnologias usadas para avaliar a Usabilidade e/ou UX de SCs;
- CI3. Publicações que discutem aspectos relacionados à avaliação de Usabilidade e/ou UX de SCs.
- Critérios para a exclusão de artigos:
 - CE1. Publicações que não atendem aos critérios acima;
 - CE2. Publicações que não têm conteúdo disponível para leitura e análise de dados (especialmente nos casos em que os trabalhos são pagos ou não estão disponíveis pelas fontes de dados);
 - CE3. Publicações que tenham um idioma diferente do inglês ou português;
 - CE4. Publicações ou arquivos que não foram revisados por pares, como relatórios técnicos, livros e trabalhos em andamento;
 - CE5. Publicações que já tenham sido adicionadas em outra fonte de dados definida em nosso MSL (duplicata).

5.1.5 Processo de seleção de artigos

A pesquisa foi realizada em fontes de dados para identificar as publicações em Agosto/2019. A seleção dos artigos foi feita pelo autor desta dissertação em conjunto com sua orientadora. A extração foi feita pelo autor dessa dissertação e revisado pela sua orientadora. De acordo com Kitchenham e Charters (2007), o uso de dois ou mais pesquisadores é essencial para manter a consistência da pesquisa e reduzir o viés. No processo de seleção preliminar (1º filtro), o primeiro pesquisador avaliou todos os trabalhos retornados com base no título e resumo e os classificou como incluídos ou excluídos com base nos critérios de inclusão e exclusão. Em seguida, a segunda pesquisadora realizou o 1º filtro separadamente. Caso houvesse

diferenças nas decisões de inclusão ou exclusão, os pesquisadores procuravam resolvê-las. Se os pesquisadores não chegassem a uma conclusão baseada apenas no título e no resumo, o trabalho era automaticamente aceito para uma leitura detalhada. Se o trabalho fosse rejeitado, uma justificativa plausível para a escolha era necessária.

No processo de seleção final (2º filtro), o primeiro pesquisador leu e classificou todos os trabalhos aprovados no 1º filtro, e extraiu os dados. Em seguida, a segunda pesquisadora verificou os artigos e justificativas de exclusão, e os artigos incluídos e suas extrações. A partir dos mesmos critérios de inclusão e exclusão, os pesquisadores selecionaram os trabalhos para que os dados fossem extraídos. Da mesma forma, se o trabalho foi excluído, foi apresentada uma justificativa plausível.

5.1.6 Definição da estratégia de extração de dados

A extração de dados deste MSL foi baseada nas respostas de cada SQ. A Tabela 5.5 mostra as possíveis respostas para cada uma das SQs. Além das respostas das SQs, foram analisados anos de publicações e locais (revista, conferência ou workshop). O protocolo completo, listas de artigos retornados e extraídos, processo de seleção e extrações estão disponíveis neste link: <https://bit.ly/31jB65q>.

Tabela 5.5: Tabela usada para a extração de dados

Referência do Artigo (Autor, Ano). Título.	
Local de publicação.	
Informações sobre a tecnologia de avaliação	
Sub-questão (SQ)	Respostas possíveis
SQ1. Qual é o critério de qualidade da tecnologia de avaliação utilizada?	A tecnologia pode ter os seguintes critérios: (i) Usabilidade : a tecnologia tem como objetivo avaliar a Usabilidade do sistema; (ii) UX : a tecnologia visa avaliar a UX do sistema; (iii) Ambos : a tecnologia visa avaliar tanto a Usabilidade quanto a UX.
SQ2. Quais aspectos da Usabilidade e/ou UX a tecnologia avalia?	As respostas obtidas no SQ2 são subjetivas e variam de artigo para artigo. Os aspectos avaliados de Usabilidade e/ou UX podem ser satisfação do usuário, eficácia, imersão, fadiga, prazer, entre outros.
SQ3. A tecnologia é específica para SCs ou para sistemas em geral?	A tecnologia pode ser: (i) Específica : A tecnologia de avaliação de Usabilidade e/ou UX é específica para SCs; (ii) Genérica : A tecnologia de avaliação de Usabilidade e/ou UX não está restrita a um tipo específico de software.

SQ4. A tecnologia foi criada pelo estudo ou é baseada em uma tecnologia já existente?	A tecnologia pode ser: (i) Existente : a tecnologia é baseada em uma tecnologia existente; (ii) Criada : a tecnologia foi criada para o estudo relatado no artigo.
SQ5. Como a tecnologia de avaliação coleta dados dos participantes?	As respostas da SQ5 são subjetivas e variam de artigo para artigo. Seu objetivo é verificar como as respostas dos usuários foram capturadas utilizando a tecnologia de avaliação, como escala Likert, perguntas abertas, múltipla escolha, <i>checklist</i> , entre outras.
SQ6. Quais são as características da tecnologia de avaliação?	As respostas da SQ6 são subjetivas e variam de artigo para artigo. O objetivo da SQ6 é extrair atributos de cada tecnologia, tais como identificar perguntas utilizadas, por exemplo, se a tecnologia é um questionário ou entrevista; identificar métricas utilizadas, por exemplo, se é uma tecnologia baseada em taxa de erro; identificar cálculos realizados, por exemplo, se é uma tecnologia de análise de eficiência; entre outros.
SQ7. A tecnologia extrai dados quantitativos ou qualitativos?	A tecnologia pode extrair dados: (i) Quantitativos : se a análise da avaliação foi feita quantitativamente; (ii) Qualitativos : se a análise da avaliação foi feita qualitativamente; (iii) Mistos : se a análise da avaliação foi feita qualitativa e quantitativamente.
SC avaliado	
SQ8. Qual é a função do SC?	Esta resposta é subjetiva e varia de artigo para artigo. Os SCs podem ser chatbots, agentes de conversação, assistentes virtuais, aplicações com funções de voz, entre outros.
SQ9. Qual categoria de aplicação do SC?	As respostas obtidas na SQ9 são subjetivas e foram identificadas durante a leitura dos trabalhos. Exemplos de categorias são saúde, educação, mobilidade urbana, facilidades diárias, entre outros.
SQ10. O sistema foi criado para um grupo específico de pessoas? Quais grupos?	As possíveis respostas são: (i) Sim : o sistema foi criado para auxiliar um grupo específico. Exemplos: cegos, surdos, idosos, crianças, entre outros; (ii) Não : o sistema não foi criado para um grupo específico.
Avaliação empírica da tecnologia de avaliação	
SQ11. A tecnologia de avaliação foi avaliada empiricamente?	As possíveis respostas são: (i) Sim : foi realizado um estudo empírico para avaliar a tecnologia de avaliação; (ii) Não : a tecnologia não foi avaliada empiricamente.

SQ11.1. Em caso afirmativo, qual foi o tipo de experimento realizado para avaliar a tecnologia de avaliação?	As respostas obtidas na SQ11.1 são subjetivas e variam de artigo para artigo. Elas podem ser estudos de viabilidade, estudos de observação, estudos de caso, entre outros.
--	--

5.1.7 Artigos selecionados após a condução do MSL

Conforme exibido na Tabela 5.6 abaixo, foram retornados 127 artigos após a aplicação da string de busca nas máquinas de busca. Um total de 52 artigos foram selecionados após a aplicação do 1º filtro, com base nos critérios de inclusão e exclusão. Um total de 39 artigos foram selecionados após a aplicação do 2º filtro (ver Apêndice C). Alguns artigos apareceram mais de uma vez em bibliotecas diferentes. Nestes casos, eles foram considerados apenas na primeira biblioteca de acordo com a ordem de busca realizada: Scopus, IEEExplore, Engineering Village e ACM.

Tabela 5.6: Total de artigos retornados e selecionados no 1º e 2º filtro.

Fonte	Total de artigos retornados	Artigos selecionados no 1º filtro	Artigos aceitos no 2º filtro
Scopus	55	39	27
IEEExplore	7	3	3
Engineering Village	51	5	4
ACM Digital Library	14	5	5
Total	127	52	39

Conforme mostrado na Tabela acima, foram extraídos os dados dos 39 artigos aprovados no 2º filtro, os quais pautaram os resultados deste MSL. Nenhum artigo extraído no primeiro MSL foi extraído novamente neste.

5.2 RESULTADOS OBTIDOS

5.2.1 Visão geral dos resultados obtidos

O resumo das resposta deste MSL estão no Apêndice F. Em relação à questão principal "*Quais as tecnologias usadas para avaliar a Usabilidade e UX dos SC?*", foram identificadas 78 tecnologias. Removendo duplicatas, ou seja, a mesma tecnologia usada em diferentes artigos, encontrou-se um total de 31 tecnologias diferentes. Estas tecnologias foram classificadas e descritas de acordo com a descrição dada pelos autores dos artigos retornados. As respostas gerais para as outras SQs são mostradas na Tabela 5.7 e explicadas nas seções posteriores. As 31 tecnologias diferentes são ilustradas na Figura 5.1 e explicadas no Apêndice D.

Tabela 5.7: Resultados gerais do segundo MSL.

SQs	Respostas Possíveis	Resultados
SQ1. Critério de Qualidade	Usabilidade	88,46%
	UX	10,26%
	Ambos	1,28%
SQ2. Aspectos	Usabilidade	89 aspectos
	UX	12 aspectos
SQ3. Especificidade da tecnologia	Genérica	51,28%
	Específica	48,72%
SQ4. Base da tecnologia de avaliação	Baseada em uma existente	60,26%
	Criada para o estudo	39,74%
SQ5. Método de coleta de dados	Respostas subjetivas	17 métodos
SQ6. Características da tecnologia de avaliação	Respostas subjetivas	-
SQ7. Tipo de análise	Quantitativa	78,21%
	Qualitativa	12,82%
	Mista	8,97%
SQ8. Função do SC	Respostas subjetivas	-
SQ9. Categoria do SC	Facilidades diárias	18 SCs
	Saúde	6 SCs
	Diversão	6 SCs
	Educação	4 SCs
	Mobilidade Urbana	2 SCs
	Computação	1 SC
	Ambiente Virtual	1 SC
	Segurança da Informação	1 SC
SQ10. Grupo de usuários	Sim	69,23%
	Não	30,77%
SQ11. Avaliação empírica	Sim	6,41%
	Não	93,59%
SQ11.1 Tipo de avaliação empírica	Estudo de viabilidade	4 tecnologias avaliadas
	Estudo de observação	1 tecnologia avaliada

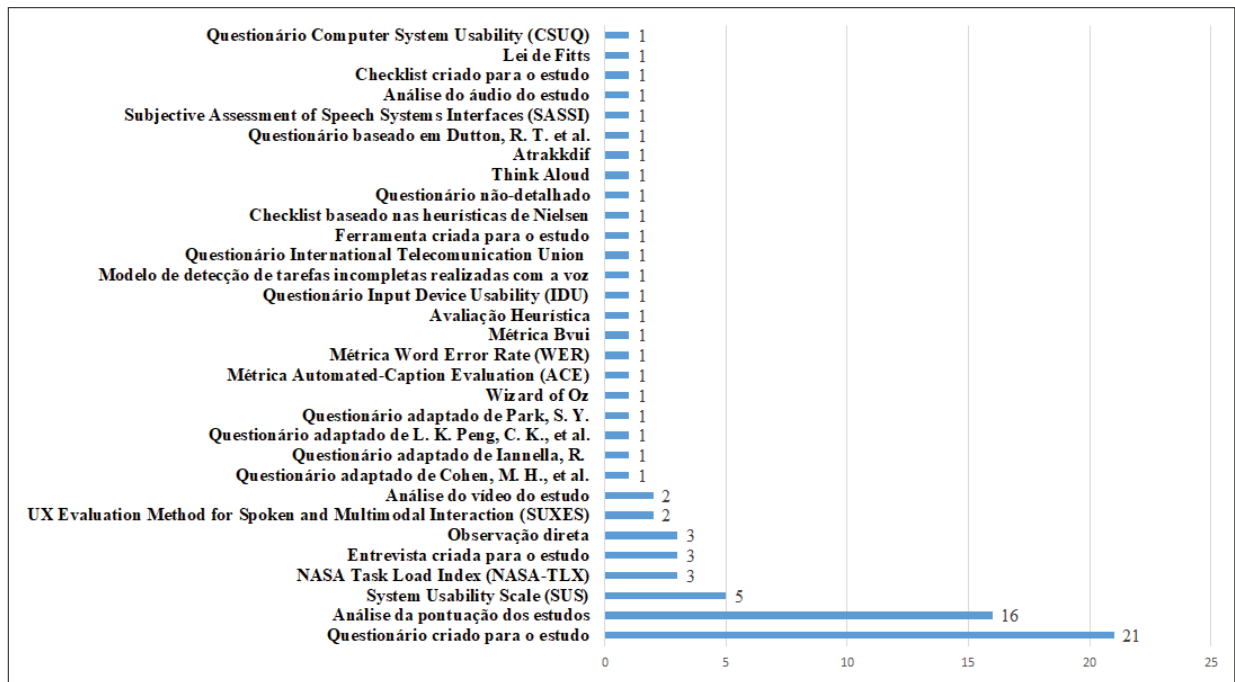


Figura 5.1: Tecnologias de avaliação utilizadas pelos autores dos artigos retornados no MSL.

5.2.2 Ano de publicação

Os artigos extraídos foram publicados entre os anos de 1999 e 2019. Cerca de 20 anos atrás, os pesquisadores já estavam pensando na qualidade do SC, e este interesse permaneceu ao longo das décadas. A Figura 5.2 ilustra a visão temporal dos trabalhos analisados. Um aumento no número de trabalhos a partir de 2008 foi observado quando as tecnologias começaram a se tornar mais acessíveis devido ao uso de *smartphones*. Até então, havia uma variação entre nenhum e poucos artigos nos anos anteriores. Em 2012, foi observada uma queda acentuada no número de artigos, mas o interesse permaneceu ao longo dos anos, expandindo-se novamente em 2017. Ainda assim, o ano de 2019 não apresenta dados para o ano inteiro, pois a busca do MSL foi realizada em Agosto. Acredita-se que, devido ao surgimento dos dispositivos assistentes controlados por voz para casas inteligentes (Amazon Echo¹ e Google Smart Home²) e o grande interesse em Inteligência Artificial, o número de trabalhos relacionados ao tema se expandirá ainda mais nos próximos anos.



Figura 5.2: Visão temporal do artigos analisados no MSL.

¹amzn.to/2R8lr3k

²https://store.google.com/?srp=/category/connected_home

5.2.3 Locais de publicação

Foram selecionados apenas trabalhos aprovados por um processo de revisão por pares. Ao todo, foram identificadas 22 conferências diferentes. A conferência com a maioria dos trabalhos analisados foi *Annual Conference of the South African Institute of Computer Scientists and Information Technologists* (SAICSIT), com dois artigos. As conferências foram organizadas com base na última classificação da *The Computing Research and Education* (CORE 2018), com destaque para *Conference on Human Factors in Computing Systems* (CHI - A*), ACM Multimedia (ACMMM - A*) e a *Annual Conference of the International Speech Communication Association* (INTERSPEECH - A).

Foram analisadas publicações de 15 revistas diferentes retornadas na busca, e foram organizadas com base na última classificação do Scimago Journal Rankings (SJR 2018). Os destaques são *IEEE Transactions on Audio, Speech, and Language Processing* (0,926) e *International Journal on Human-Computer Studies* (0.688). Foi extraído um trabalho publicado em workshop, no *Spoken Language Technology Workshop*. A próxima seção mostra os resultados de cada uma das SQs.

5.2.4 SQ1. Critério de Qualidade

Os resultados da SQ1 indicaram que 88,46% das tecnologias concentraram sua avaliação no critério de Usabilidade. Em Amith et al. (2019), os autores avaliaram a facilidade de uso, eficiência e capacidades esperadas de um agente de conversação que alertava os usuários sobre vacinas. A tecnologia utilizada foi um questionário baseado em Cohen et al. (2004), que continha 3 perguntas a serem respondidas por uma escala Likert de 7 pontos que variava entre "discordo fortemente" e "concordo fortemente".

Cerca de 10,26% das tecnologias foram utilizadas para avaliar a UX dos SCs. Em Gürkök et al. (2017), a tecnologia SUXES (Turunen et al., 2009) foi usada para avaliar aspectos de UX, tais como naturalidade, prazer e fadiga. Além disso, 1,28% das tecnologias identificadas foram utilizadas para avaliar tanto a Usabilidade quanto a UX. No único caso retornado, Moustakas et al. (2011) desenvolveram uma entrevista com perguntas abertas para o estudo, onde avaliaram o uso da modalidade, a facilidade de uso, a satisfação do usuário e a UX em geral.

Esses resultados evidenciam a falta de tecnologias que avaliem a Usabilidade e a UX em conjunto. A maioria das tecnologias avaliam aspectos pragmáticos da Usabilidade. Algumas avaliam os aspectos hedônicos da UX. No entanto, a menor parte avalia os dois critérios em conjunto. A avaliação conjunta da Usabilidade e UX pode ser um passo que impacta a qualidade do sistema, permitindo que os objetivos comportamentais sejam avaliados simultaneamente com os sentimentos hedônicos dos usuários.

5.2.5 SQ2. Aspectos de Usabilidade e UX

Os resultados obtidos nessa subquestão identificaram 89 aspectos diferentes utilizados para avaliar a Usabilidade e 12 aspectos diferentes para avaliar a UX no contexto de software de voz. Para reduzir o viés dos pesquisadores, os aspectos foram classificados de acordo com a definição dada pelos autores dos artigos. Sendo assim, foi identificado que alguns aspectos, por exemplo eficácia, facilidade de uso e utilidade, são considerados em alguns casos como um aspecto de Usabilidade e em outros como UX. A Figura 5.3 mostra os aspectos de Usabilidade e a Figura 5.4 os de UX.

Um exemplo de aspecto de Usabilidade pode ser encontrado no trabalho apresentado por Miguel-Hurtado *et al.* (2016), onde os autores calcularam a eficiência a partir do tempo gasto pelos participantes para executar as tarefas após aprenderem a utilizar um sistema de autenticação baseado em voz.

Um exemplo de aspecto de UX é mostrado em Moustakas *et al.* (2011), onde os autores utilizaram a observação direta para avaliar a UX geral do usuário ao utilizar um jogo controlado por voz. Neste caso, dois pesquisadores anotavam as observações enquanto os usuários realizavam as tarefas.

Os resultados da SQ2 vão em direção à afirmativa de que há múltiplas definições do que é Usabilidade e UX no contexto de software de voz, dada por Kocaballi *et al.* (2018). Isso porque alguns aspectos, como eficácia e facilidade de uso, são tratados por alguns autores como Usabilidade e por outros como UX. Ainda, é possível constatar que os principais aspectos de Usabilidade definidos pela ISO 9241-11 (1998) são avaliados também no contexto de voz, que são a eficácia, eficiência e satisfação do usuário. Já para o contexto de UX, não é possível identificar aspectos que se destacam, e os autores preferem avaliar a UX de forma genérica.

5.2.6 SQ3. Especificidade da tecnologia de avaliação

Cerca de 51,28% das tecnologias retornadas não são específicas para os softwares de interação por voz. Ou seja, podem avaliar qualquer tipo de software, tanto os de interação por voz, como os de *touchscreen*, sistemas *web*, entre outros. Por exemplo, Tchankue *et al.* (2012) utilizam o SUS (Brooke, 1996), uma tecnologia de avaliação utilizada em diversos contextos, para avaliar a Usabilidade de um protótipo de sistema de comunicação de carros. O SUS contém 10 afirmações sobre Usabilidade em geral que podem ser respondidas por uma escala Likert de 5 pontos, que varia entre "discordo fortemente" e "concordo fortemente".

Cerca de 48,72% das tecnologias retornadas foram específicas para o contexto da voz. Por exemplo, Kafle e Huenerfauth (2017) utilizaram duas métricas de reconhecimento automatizado de fala (*Automated-Caption Evaluation* - ACE e *Word Error Rate* - WER) para calcular a eficácia de um gerador de legenda automático. As métricas ACE e WER são duas tecnologias que, por meio de cálculos com diversas variáveis (comprimento do texto de referência,

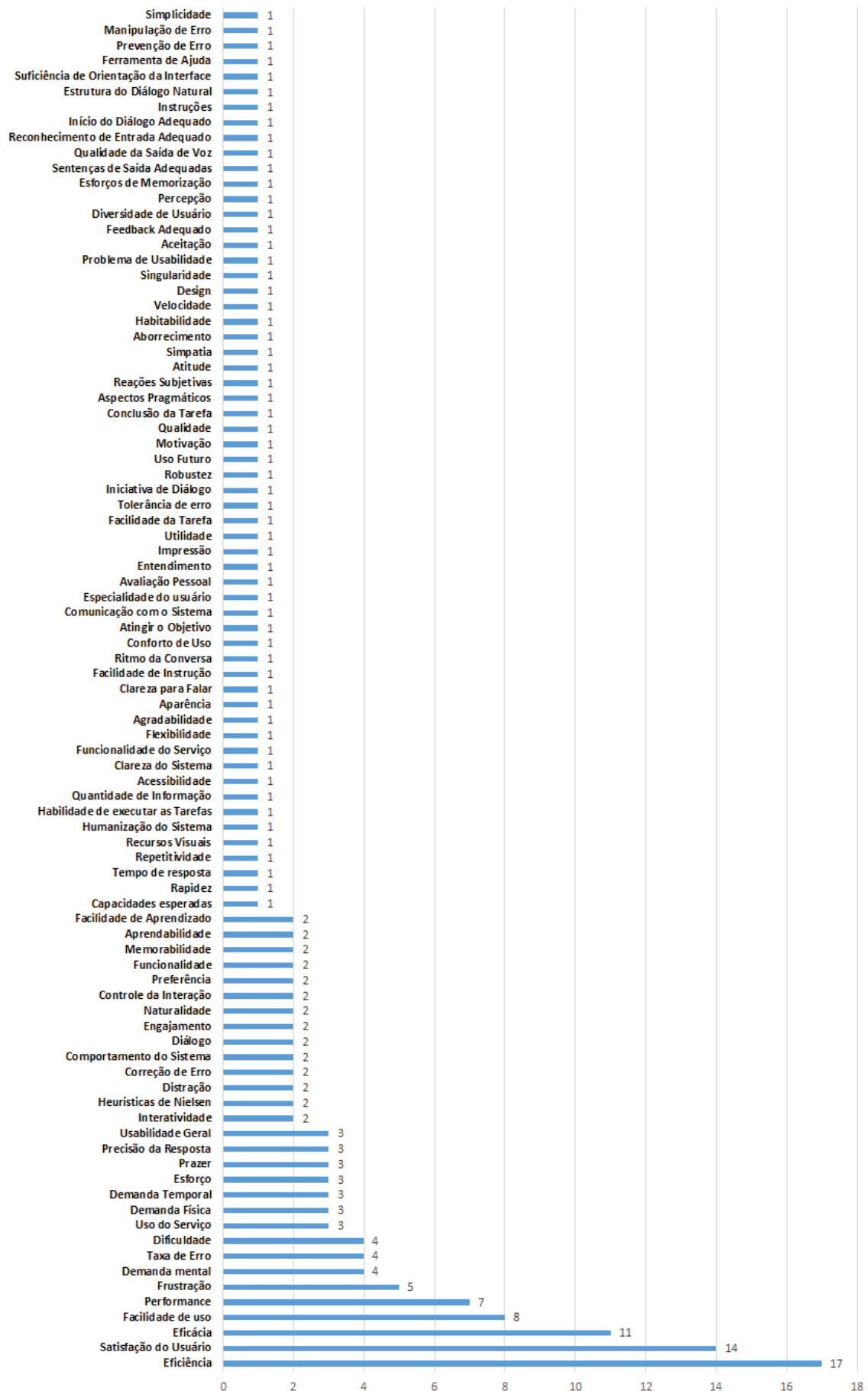


Figura 5.3: Aspectos de Usabilidade avaliados pelas tecnologias.

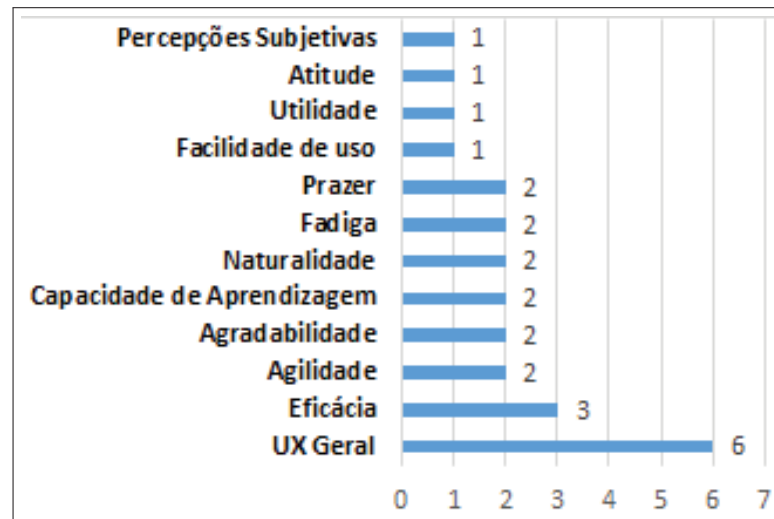


Figura 5.4: Aspectos de UX avaliados pelas tecnologias.

número de erros, número de palavras corretas, entre outras), permitem que os avaliadores descubram se um gerador de legenda que reconhece a fala de forma automatizada é eficaz.

Os resultados da SQ3 mostram um certo equilíbrio na utilização de tecnologias específicas e genéricas. O uso de tecnologias genéricas possui um lado positivo relacionado a reutilização, ou seja, a possibilidade de utilizar a mesma tecnologia em diversos contextos. No entanto, no caso de sistemas de interação baseado em voz, acredita-se que seja necessário que a tecnologia possua direcionamentos específicos durante a avaliação, visto que a interação por voz possui suas especificidades que a difere de interações convencionais, como o ritmo do diálogo entre o usuário e o sistema de voz. As tecnologias específicas de avaliação de interação por voz atingem esta característica de especificidade, que impacta positivamente na qualidade do software durante o processo de avaliação.

5.2.7 SQ4. Base da tecnologia de avaliação

Os resultados dessa subquestão mostram que 60,26% das tecnologias utilizadas já existem e foram utilizadas nos estudos relatados nos artigos. Um exemplo de tecnologia existente pode ser encontrado no trabalho de Neto et al. (2009), o qual os autores utilizam um *checklist* baseados nas Heurísticas de Nielsen (Nielsen, 1993) para avaliar a Usabilidade. Para cada heurística, os autores criaram uma sentença em forma de *checklist* com o objetivo de guiar os avaliadores na busca de problemas no software.

Em contrapartida, 39,74% das tecnologias retornadas foram criadas especificamente para o estudo relatado no artigo. Por exemplo, no trabalho de Huerta et al. (2011), os autores criaram o seu próprio questionário para avaliar um software desenvolvido para apoiar o processo de inovação e ideia a partir do reconhecimento de fala. Este questionário continha 7 sentenças sobre Usabilidade geral que podiam ser respondidas por uma escala Likert de 4 pontos que

variava entre "discordo totalmente" e "concordo totalmente", além de um espaço para comentário sobre a resposta.

Em relação aos resultados da SQ4, pode-se notar que grande parte dos trabalhos se baseiam em uma tecnologia existente na literatura. Reforçando a análise mencionada na subquestão anterior, a reutilização é uma característica presente nas tecnologias de avaliação de sistemas de interação por voz. No entanto, ainda há as tecnologias que são criadas para o estudo, que, por sua vez, consideram as especificidades do sistemas de interação por voz, mas geralmente se limitam a um único estudo.

5.2.8 SQ5. Método de coleta das respostas dos participantes

Os resultados dessa subquestão mostram que foram identificadas 17 métodos de coleta de dados, as quais são exibidas na Figura 5.5. O método de coleta mais utilizado é a escala Likert, utilizada em 41 ocasiões, onde os usuários responderam uma escala de concordância de acordo com a afirmação/pergunta fornecida pelos pesquisadores. Essa escala foi utilizada por Wenceslao e Estuar (2019), onde os autores coletaram respostas a partir de 10 afirmações que podiam ser respondidas em uma escala de 5 pontos, que varia entre discordo fortemente e concordo fortemente.

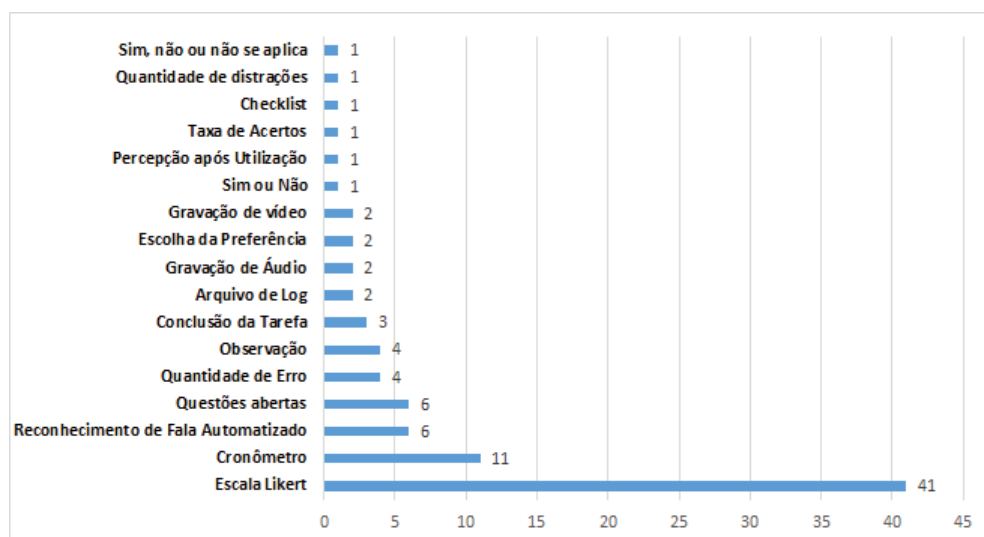


Figura 5.5: Tipo de coleta de dados utilizado pelas tecnologias.

Outros métodos de coleta de dados se destacaram, como o cronômetro (análise a partir do tempo gasto), o reconhecimento de fala automatizado (análise do fala do usuário) e as questões abertas (análise qualitativa das respostas dos usuários).

De acordo com os resultados da SQ5, é possível verificar que a escala Likert foi amplamente utilizada pelas tecnologias de avaliação. A alta utilização dos questionários, tanto os existentes quanto os criados para o estudo, contribuem para que essa escala seja o método de coleta mais utilizado. É possível perceber também que tanto essa escala quanto o cronômetro, segundo método de coleta mais utilizado, contribuem para que os resultados da SQ7 sejam em sua

maioria quantitativos, visto que é este o tipo de análise que esses métodos de coleta possibilitam. Os métodos de coleta que fornecem análise qualitativa estão presentes nas tecnologias, por exemplo questões abertas e observação, mas fazem parte do grupo das minorias. Entretanto, uma análise qualitativa é de grande importância para a avaliação da qualidade do software, visto que permite identificar contribuições subjetivas e descritivas que podem auxiliar na melhoria e evolução dos sistemas.

5.2.9 SQ6. Características das tecnologias de avaliação

Para a tecnologia mais utilizada (Questionário criado para o estudo), esta questão extraiu todas as perguntas/afirmações contidas nos questionários. Por exemplo em Azeta et al. (2018), os autores criaram um questionário com 17 afirmações para avaliar uma tecnologia assistiva baseada em voz para auxiliar estudantes com deficiência visual a responder exames em plataformas EAD. Entre as afirmações, pode-se citar como exemplo "*O sistema trabalhou da forma como eu esperei na conversação*" e "*O ritmo da interação com o sistema foi apropriado*".

No caso da segunda tecnologia de avaliação mais utilizada, a análise das pontuações do estudo, a SQ6 extraiu informações de como foi feita a análise e com base em quais pontuações obtidas no estudo. Por exemplo em Chang et al. (2009), a análise foi feita a partir do número de participantes que não completaram a tarefa e a média do tempo para a realização das tarefas.

Cada tipo de tecnologia teve suas características extraídas de alguma forma. Das tecnologias de questionário e entrevista, foram extraídas as questões e afirmações utilizadas. Das tecnologias de análise, foram extraídos os métodos e os dados utilizados. Das tecnologias de métrica e modelo, foi extraído o cálculo realizado. A descrição de cada tecnologia pode ser vista no Apêndice D.

Com a SQ6 foi possível verificar as características de cada tecnologia de avaliação. Os questionários criados para o estudo apresentam questões desenvolvidas pelos próprios pesquisadores envolvidos no estudo, e o número de questões varia de acordo com a necessidade de cada um. A análise das pontuações do estudo permitem utilizar os dados recolhidos durante a realização das tarefas para analisar posteriormente de acordo com os cálculos definidos. É possível identificar uma autonomia dos pesquisadores ao criarem suas próprias tecnologias de avaliação, cálculos e métricas. Da mesma forma, pode-se observar uma necessidade dos pesquisadores com a criação de muitos questionários que avaliam apenas um estudo.

5.2.10 SQ7. Tipo de análise

Os resultados dessa subquestão revelam que 78,21% das tecnologias avaliaram os softwares de voz apenas quantitativamente. Por exemplo em Wargnier et al. (2018), os autores realizaram a análise quantitativa a partir das respostas obtidas com a escala Likert de 4 pontos de um questionário. Esta escala variava entre "muito negativo" e "muito positivo", as quais são

as respostas para as 5 questões sobre agradabilidade, aparência, clareza para fala, facilidade de instrução e ritmo da conversa.

Cerca 12,82% das tecnologias avaliaram os SCs de forma qualitativa. É o caso de Bernhaupt et al. (2007), onde os autores analisaram o áudio do estudo para realizar uma análise qualitativa a partir dos *insights* obtidos com o *think aloud*. Esse método visa identificar possíveis pensamentos dos usuários a partir do que eles falam em voz alta durante o experimento.

Apenas 8,97% das tecnologias avaliaram os softwares tanto de forma quantitativa como qualitativa. Um exemplo pode ser visto em Hara et al. (2010b), onde os autores criaram um questionário que continha questões para serem respondidas com escala Likert, questão para atribuir um valor variante entre 0 e 100 e questões abertas para captar respostas dissertativas para uma posterior análise qualitativa.

Os resultados da SQ7 estendem a análise feita dos resultados da SQ5. O tipo de análise mais utilizado é o quantitativo, enquanto o qualitativo e o "em conjunto" (quantitativo e qualitativo) fazem parte de uma pequena parcela das tecnologias de avaliação. Com isso, é possível inferir que a maioria dos autores preferem focar suas análises nos números obtidos pelas tecnologias de avaliação e não consideram as subjetividades de um pensamento do usuário que pode estar inserido em um comentário qualitativo, por exemplo. Entretanto, uma tecnologia de avaliação que integre tanto respostas quantitativas quanto qualitativas pode fornecer aos pesquisadores um conjunto de dados mais diversificado para que seja realizada a análise conforme as suas possibilidades.

5.2.11 SQ8. Função do SC

Esta subquestão extraiu dados subjetivos em relação aos SCs. Ela foi utilizada para captar uma breve descrição das funcionalidades dos softwares avaliados. Por exemplo em Wenceslao e Estuar (2019), os autores avaliaram um registro médico eletrônico, onde o sistema utilizava o reconhecimento de fala para identificar padrões de sintomas mencionados por um paciente durante uma consulta. Já em Amith et al. (2019), o software avaliado foi um agente conversacional utilizado para aconselhar os usuários sobre a vacina do vírus do papiloma humano (*Human Papilloma Virus* - HPV).

Os resultados da SQ8 permitem perceber que os sistemas de voz desempenham diversas funções. Desde agentes conversacionais para aconselhamento dos usuários sobre a prevenção de doenças, até gerador de legenda automático baseado no reconhecimento de fala. Os resultados desta SQ mostram a presença dos softwares de voz em diversas áreas da sociedade, que permanece ao longo dos anos e contribui de alguma forma com os usuários que os utilizam.

5.2.12 SQ9. Categoria do SC

Esta SQ buscou enquadrar os softwares identificados na subquestão SQ8 em categorias que foram sendo identificadas durante a análise dos estudos. A Figura 5.6 mostra as categorias identificadas.

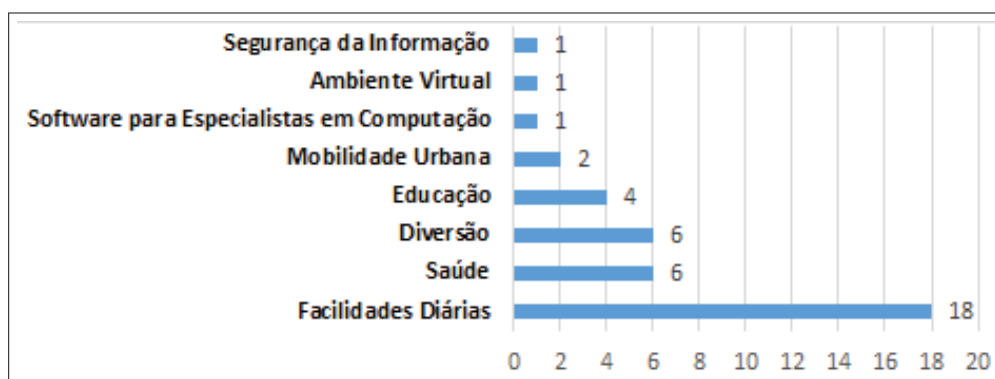


Figura 5.6: Categorias dos SCs identificados nos artigos.

Um exemplo de artigo que aborda a categoria de facilidades diárias é Hara et al. (2010a), onde os autores avaliaram um sistema onde o usuário podia procurar e colocar música para tocar no computador por meio da interação por voz. Um exemplo de software ligado à saúde pode ser encontrado em Kusumaningayu e Ayu (2017), onde os autores avaliaram uma ferramenta de acessibilidade para web que auxilia pessoas cegas e deficientes visuais a acessarem a internet. Já em Farinazzo et al. (2010), foi avaliado um software educativo que tinha como objetivo melhorar a pronúncia do usuário em relação ao inglês.

De acordo com os resultados desta SQ, pode-se inferir que um dos principais objetivos dos SCs é facilitar o cotidiano de quem a utiliza. Isso pode ser exemplificado com a expansão da utilização dos assistentes virtuais, agentes conversacionais e até mesmo com a criação de dispositivos físicos para assistência virtual, como os dispositivos Amazon Echo e Google Smart Home, os quais tem como função principal fornecer assistências diárias para seus usuários. Também é possível verificar a utilização da interação por voz na saúde, educação, diversão e outras áreas, o que nos faz perceber que a interação por voz não se restringe apenas às facilidades gerais, mas em alguns casos possui motivação específica inserida.

5.2.13 SQ10. Grupo de usuários de SC

Os resultados obtidos com essa subquestão mostram que 69,23% dos softwares avaliados não foram criados para um grupo específico de pessoas. Em contrapartida, 30,77% dos softwares foram criados com o objetivo de auxiliar algum grupo. Os grupos que foram identificados são mostrados na Figura 5.7.

Como pode ser observado, o grupo que obteve maior direcionamento dos softwares de voz retornados foi o de deficientes visuais. Um exemplo desse auxílio pode ser visto em Kusumaningayu e Ayu (2017), que avaliaram uma ferramenta de acessibilidade baseada em



Figura 5.7: Grupos de pessoas que os softwares de voz buscaram auxiliar.

voz que auxilia pessoas cegas a navegarem na web. Em seguida, idosos e deficientes auditivos também se destacaram quando o software de voz buscou auxiliar algum grupo específico de pessoas.

Apesar da maioria dos SCs analisados não serem criados para auxiliar um grupo de pessoas, os resultados da SQ10 mostram que alguns possuem esse direcionamento. Analisando esses grupos, pode ser observada uma certa ligação dos SCs com a acessibilidade e inclusão digital, visto que, quando direcionados para auxiliar algum grupo, os SCs focam em pessoas que podem possuir algum tipo de dificuldade ao utilizar interações convencionais, como deficientes visuais, surdos e idosos.

5.2.14 SQ11. Avaliação empírica da tecnologia de avaliação

Os resultados desta subquestão mostram que apenas 6,41% das tecnologias de avaliação (5 tecnologias) foram avaliadas empiricamente. Um exemplo de avaliação empírica está descrito em Hone e Graham (2000), onde os autores realizaram diversos estudos para lapidar e evoluir a tecnologia *Subjective Assessment of Speech Systems Interfaces* (SASSI) que eles propuseram.

Dentro desta subquestão, há a SQ11.1 (tipo de avaliação empírica). Os resultados da SQ11.1 mostram que das 5 tecnologias que passaram por uma avaliação empírica, 4 foram por estudo de viabilidade, como é o caso do SASSI, citado no exemplo anterior. Apenas a métrica *Bvui*, criada por Schaffer et al. (2015) para calcular a eficiência, passou por um estudo de observação.

Esses resultados mostram que muitas tecnologias são utilizadas sem antes passar por uma avaliação empírica. Isso é embasado com a alta utilização dos questionários criados para um estudo apenas. De acordo com os resultados, poucas tecnologias de avaliação passam pelo processo importante de empirismo, que busca a validação, verificação de melhorias, verificação de viabilidade, entre outros aspectos. Além disso, pode-se inferir que os pesquisadores, por conta da escassez citada anteriormente somada à demanda de tempo gasto para realizar essas avaliações empíricas, preferem criar os seus próprios questionários e utilizarem sem avaliações.

5.3 SÍNTESE DO CAPÍTULO

Este capítulo apresentou um MSL executado para caracterizar as tecnologias de avaliação de Usabilidade e/ou UX que estão sendo utilizadas para avaliar e melhorar a qualidade dos SCs. De um total inicial de 127 artigos retornados, 39 foram selecionados para extração. A partir destes artigos, foram encontradas 31 tecnologias de avaliação diferentes que tiveram, a partir das subquestões, as suas características extraídas.

Os resultados mostram que os pesquisadores estão criando os seus próprios questionários de avaliação, específicos para o seu estudo, e que não passam por nenhum processo de validação ou avaliação empírica. Além disso, pode-se notar também avaliações apenas de Usabilidade, enquanto a UX ainda não possui aspectos bem definidos neste contexto. Foi identificada também a utilização de tecnologias genéricas, as quais não levam em consideração as especificidades que um SC possui. Ainda, a maioria das tecnologias realizam apenas análise quantitativa, que não leva em consideração possíveis pensamentos subjetivos dos usuários que podem contribuir com a qualidade do software.

Em relação aos SCs, foi visto que eles são amplamente utilizados para auxiliar no cotidiano da vida das pessoas, mas também estão sendo estudados em contextos específicos como saúde e educação. Foi verificado também uma possível correlação dos SCs com questões de acessibilidade e inclusão digital, visto que, quando criados para auxiliar algum grupo específico, estes softwares são focados em pessoas que possuem alguma dificuldade, como deficientes visuais ou auditivos, idosos e pessoas com *Alzheimer*.

Sumarizando os resultados, as lacunas identificadas foram:

- Falta de tecnologias de avaliação para SCs que consideram em conjunto os conceitos de Usabilidade e UX;
- A maioria das tecnologias de avaliação utilizadas não passam por um processo de avaliação empírica e validação;
- Há a utilização de tecnologias de avaliação genéricas que não consideram as especificidades que um SC possui;
- As tecnologias de avaliação geralmente focam apenas na análise quantitativa.

6 PROPOSTA DA SEGUNDA VERSÃO DA U2XECS

A partir dos resultados do estudo exploratório e do segundo MSL, a U2XECS foi evoluída para sua segunda versão. Este capítulo apresenta as alterações feitas e o template da segunda versão da U2XECS.

6.1 ALTERAÇÕES REALIZADAS

Para propor a evolução da U2XECS, primeiro foram abordados os resultados do estudo exploratório. As negações presentes nas afirmativas da U2XECS foram retiradas do meio da afirmativa e alocadas no início. A palavra "efeito", da afirmativa 5, foi melhor elaborada. Além disso, a palavra "maçante" foi substituída. A Tabela 6.1 mostra as alterações realizadas com base no estudo exploratório.

Tabela 6.1: Alterações realizadas com base no estudo exploratório.

Primeira Versão da U2XECS	Segunda Versão da U2XECS
A interação por voz me deixou desconfiante para usar esse sistema.	Não me senti confiante ao utilizar a voz para interagir com este sistema.
A interação por voz causa efeito imediato no sistema.	O sistema respondeu de maneira imediata a minha interação por voz.
Foi muito maçante utilizar a voz para realizar essas tarefas.	Foi muito entediante utilizar a voz para realizar essas tarefas no sistema.

Uma das recomendações também do estudo exploratório foi que se adicionasse mais afirmativas no questionário. Dessa forma, foi feita uma análise com base nas tecnologias de avaliação retornadas no segundo MSL, descritas no Apêndice D, para verificar os pontos e afirmativas que poderiam ser adicionadas na U2XECS. Sendo assim, foram adicionadas 18 novas afirmativas com base na análise dos resultados do MSL. A Tabela 6.2 mostra as afirmativas adicionadas e de onde foram extraídas (artigos retornados no segundo MSL).

Além das alterações provindas do estudo exploratório e do segundo MSL, também foi realizada: (i) atualização do tempo verbal de todas as afirmativas (para o passado); (ii) simplificação ou reformulação de algumas afirmativas, mas mantendo o mesmo sentido; (iii) adição de uma questão qualitativa sobre cada aspecto. O template da segunda versão da U2XECS é mostrado na próxima seção.

Tabela 6.2: Adição de 18 afirmativas com base nos resultados do MSL.

Afirmativa	Fontes
Consegui me familiarizar com o sistema de interação por voz quando o usei pela primeira vez.	(Azeta et al., 2018)
O sistema se comportou da forma como eu esperei durante a interação por voz.	(Azeta et al., 2018; Fernandez Martinez et al., 2008)
Eu achei fácil entender como interagir por voz no sistema.	(Fernandez Martinez et al., 2008)
Foi fácil se tornar habilidoso para usar o sistema conversacional.	(Kusumaningayu e Ayu, 2017)
O ritmo da interação por voz com o sistema foi apropriado.	(Azeta et al., 2018)
Durante a minha interação por voz com o sistema tive que repetir diversos comandos.	(Spiliotopoulos et al., 2009)
Consegui completar minha tarefa com a interação por voz em um tempo que acredito ser razoável.	(Azeta et al., 2018)
O sistema conversacional foi capaz de se recuperar facilmente de algum erro ou equívoco que eu cometi.	(Azeta et al., 2018; Gürkök et al., 2011, 2017)
Eu completei minhas tarefas com interação por voz no sistema sem cometer erros.	(Gürkök et al., 2011, 2017)
O sistema conversacional me forçou usar palavras-chave.	(Farinazzo et al., 2010)
Quando a entrada de dados no sistema conversacional foi inconsistente ou ambígua, o sistema solicitava mais informações.	(Farinazzo et al., 2010)
Do ponto de vista dos usuários novatos, o sistema conduziu, de uma boa maneira, a interação por voz.	(Farinazzo et al., 2010)
Do ponto de vista dos usuários avançados, o sistema permitiu uma grande quantidade de dados de entrada de uma só vez.	(Farinazzo et al., 2010)
Nunca tive uma experiência de interação por voz como tive ao realizar estas tarefas no sistema.	(Moustakas et al., 2011)
O sistema respondeu a minha interação por voz de uma maneira amigável.	(Azeta et al., 2018)
Eu me diverti ao utilizar a voz para realizar essas tarefas no sistema.	(Gürkök et al., 2011, 2017)
Eu me senti no controle do sistema durante a interação por voz.	(Farinazzo et al., 2010; Spiliotopoulos et al., 2009)
Eu acho que usaria a interação por voz regularmente nesse sistema ao invés de outro tipo de interação.	(Fernandez Martinez et al., 2008)

6.2 TEMPLATE DO QUESTIONÁRIO DA SEGUNDA VERSÃO DA U2XECS

QUESTIONÁRIO DA U2XECS
Nome:
<p>– Por favor, preencha o questionário abaixo para avaliar a sua experiência na interação com o sistema conversacional. Sua resposta é importante e nos ajudará a propor melhorias por meio desta avaliação.</p> <p>– Nas questões 1 a 36, assinale apenas uma resposta por questão. Nas questões marcadas com "*" e da 37 a 39, escreva o quanto desejar. Se necessário, utilize o verso dessas folhas.</p> <p>– Não há respostas certas ou erradas. O importante para nós é a sua opinião.</p> <p>– Responda as questões de múltipla escolha utilizando a seguinte escala:</p> <p>1 = discordo totalmente;</p> <p>2 = discordo parcialmente;</p> <p>3 = não discordo nem concordo;</p> <p>4 = concordo parcialmente;</p> <p>5 = concordo totalmente.</p>

Satisfação do Usuário					
	1	2	3	4	5
1. Foi fácil utilizar a voz para realizar as tarefas nesse sistema.					
2. Precisei aprender muito sobre o sistema antes de realizar essas tarefas com a voz.					
3. Consegui me familiarizar com o sistema conversacional quando o usei pela primeira vez.					
4. Eu me senti satisfeito ao usar a voz para realizar essas tarefas.					
5. O sistema se comportou da forma como eu esperei durante a interação por voz.					
6. Eu achei fácil entender como interagir por voz no sistema.					
7. Foi fácil se tornar habilidoso ao utilizar o sistema conversacional.					
* Descreva os problemas de satisfação que você identificou no sistema.					
Eficiência					
	1	2	3	4	5
8. O sistema respondeu de maneira imediata a minha interação por voz.					
9. Eu usaria a voz com frequência para realizar essas tarefas no sistema.					
10. O ritmo da interação por voz com o sistema foi apropriado.					

11. Durante a minha interação por voz com o sistema tive que repetir diversos comandos.					
12. Consegui completar minha tarefa com a interação por voz em um tempo que acredito ser razoável.					
* Descreva os problemas de eficiência que você identificou no sistema.					
Eficácia					
	1	2	3	4	5
13. O sistema não reconheceu o que eu falei durante a minha interação por voz.					
14. A minha interação por voz não foi útil para realizar as tarefas no sistema.					
15. O sistema conversacional foi capaz de se recuperar facilmente de algum erro ou equívoco que eu cometi.					
16. Eu completei minhas tarefas com interação por voz no sistema sem cometer erros					
17. O sistema conversacional me obrigou a usar palavras-chave.					
18. Quando a entrada de dados no sistema conversacional foi inconsistente ou ambígua, o sistema solicitava mais informações.					
19. Do ponto de vista dos usuários novatos, o sistema conduziu, de uma boa maneira, a interação por voz.					
20. Do ponto de vista dos usuários avançados, o sistema permitiu uma grande quantidade de dados de entrada de uma só vez.					
* Descreva os problemas de eficácia que você identificou no sistema.					
UX Genérica					
	1	2	3	4	5
21. Realizar essas tarefas com a voz no sistema foi uma experiência boa.					
22. Nunca tive uma experiência de interação por voz como tive ao realizar estas tarefas no sistema.					
* Descreva os problemas de UX que você identificou no sistema.					
Afeto/Emoção					
	1	2	3	4	
23. Me senti desanimado ao utilizar a voz para realizar essas tarefas no sistema.					
24. Não me senti confiante ao utilizar a voz para interagir com este sistema.					
25. O sistema respondeu a minha interação por voz de uma maneira amigável.					
* Descreva os problemas de afeto e emoção que você identificou no sistema.					

Prazer/Diversão					
	1	2	3	4	5
26. Foi muito prazeroso utilizar a voz para realizar essas tarefas no sistema.					
27. Foi muito entediante utilizar a voz para realizar essas tarefas no sistema.					
28. Eu me diverti ao utilizar a voz para realizar essas tarefas no sistema.					
* Descreva os problemas de prazer e diversão que você identificou no sistema.					
Estética/Atração					
	1	2	3	4	5
29. O sistema possuía um design inovador que facilitou a realização das tarefas por meio da voz.					
30. Utilizar a voz para realizar as tarefas me atraiu para usar o sistema.					
* Descreva os problemas de estética e atração que você identificou no sistema.					
Engajamento/Fluxo					
	1	2	3	4	5
31. Utilizar a voz fez com que eu me dedicasse mais para realizar as tarefas no sistema.					
32. Eu não consegui me concentrar na realização das tarefas ao utilizar a interação por voz.					
33. Eu me senti no controle do sistema durante a interação de voz.					
* Descreva os problemas de engajamento e fluxo que você identificou no sistema.					
Motivação					
	1	2	3	4	5
34. Me senti motivado ao utilizar a voz para realizar essas tarefas no sistema.					
35. Eu acho que usaria a interação por voz regularmente nesse sistema ao invés de outro tipo de interação.					
* Descreva os problemas de motivação que você identificou no sistema.					
Encantamento					
	1	2	3	4	5
36. As tarefas que realizei no sistema com a voz fez com que eu ficasse encantado com este tipo de interação.					
* Descreva os problemas de encantamento que você identificou no sistema.					

37 - No geral, utilizar a voz para realizar as tarefas foi _____, porque:

38 - No geral, um ponto **positivo** ao utilizar a interação por voz para realizar essas tarefas é _____, porque:

39 - No geral, um ponto **negativo** ao utilizar a interação por voz para realizar essas tarefas é _____, porque:

7 SURVEY EXPLORATÓRIO

A partir da proposição da segunda versão da U2XECS, foi realizado um survey com usuários e ex-usuários de SCs, com o objetivo de verificar suas opiniões, identificar insights sobre o uso geral dos SCs e possíveis dificuldades de uso no contexto brasileiro. Dessa forma, os resultados do survey auxiliaram na evolução da tecnologia U2XECS para sua terceira versão.

7.1 PLANEJAMENTO E EXECUÇÃO

O alvo de nosso survey eram os brasileiros que usam ou usavam um SC. Nosso objetivo era identificar as opiniões das pessoas de qualquer área de conhecimento, sexo, idade ou nível de educação sobre SCs no contexto brasileiro.

Nosso survey seguiu as recomendações de Lazar et al. (2017) para desenvolver as perguntas abertas e fechadas, estruturar a survey e realizar o teste piloto. O survey foi preparado e aplicado em português, por ser a língua nativa do autor. As perguntas utilizadas no survey foram:

1. *Qual(is) sistema(s) conversacional(is) você usa ou usou?* [Por exemplo: Siri, Cortana, Alexa, Google Assistant, entre outros];
2. *Qual tipo de atividade você realiza(ou) com o sistema conversacional?* [Por exemplo, atividades diárias, aprendizagem, entretenimento, outros];
3. *Atualmente, com que frequência você usa sua voz como forma de interação em sua vida cotidiana?* [Escala Likert de 5 pontos variando entre "Eu não uso mais" até "Eu uso diariamente"];
4. *Se você usava um sistema conversacional e não usa mais, poderia descrever as razões para isso?* [Resposta aberta];
5. *Como você avalia o uso do sistema conversacional?* [Escala Likert de 5 pontos variando entre "Fácil" a "Difícil"];
6. *Você poderia descrever alguma dificuldade que você identificou ao utilizar o sistema conversacional? Se você não teve nenhuma dificuldade, poderia explicar por que?* [Resposta aberta];
7. *Quando você usa uma função de voz no sistema, você geralmente:* consegue fazer o que quer imediatamente; consegue fazer o que quer, mas tem que repetir a ação algumas vezes; ou consegue fazer apenas procurando por tutoriais];

8. *Do seu ponto de vista, você acha que os sistemas conversacionais são intuitivos? Por exemplo, você sempre sabe o que fazer no sistema.* [Escala Likert de 5 pontos variando entre "Discordo totalmente" a "Concordo totalmente"].;
9. *Quando você tem outra opção de interação no software (além da voz), você:* [usa a outra opção primeiro mas tenta usar a voz em outras ocasiões; usa a voz primeiro mas tenta usar a outra opção em outras ocasiões; usa somente a outra opção; ou usa somente a voz];
10. *Em geral, como você avalia o uso do sistema conversacional?* [Escala Likert de 5 pontos variando entre "Muito frustrante" a "Muito satisfatório"].;
11. *Cite pelo menos uma característica positiva do sistema conversacional.*
12. *Cite pelo menos uma característica negativa do sistema conversacional.*

Um teste piloto foi realizado com cinco pessoas para verificar se o questionário era claro e conciso, sendo uma pessoa da área de IHC, uma da área de saúde, uma da área de ciências agrárias, uma da área de informática na educação e uma da área de contabilidade. Após o feedback dos participantes do teste piloto, foram sugeridas modificações no survey, tais como melhoria na clareza das perguntas e mudanças de alguns termos.

Foi utilizado o Google Forms para tornar o survey disponível online. O link foi compartilhado em vários grupos de todas as áreas do conhecimento, convidando as pessoas a participar do survey. Todos os respondentes marcaram "Sim" no Termo de Consentimento apresentado no início do survey, concordando com sua participação no fornecimento de dados para nossa pesquisa. Além disso, eles responderam perguntas sobre áreas de ocupação, nível de educação, gênero e idade. Foram obtidas 114 respostas. No entanto, um respondente não forneceu informações sobre o SC utilizado por ele; portanto, as respostas deste respondente foram excluídas. Dessa forma, foram analisados dados de 113 respondentes.

Para analisar os dados quantitativos, foram gerados gráficos e tabelas sobre as respostas dos participantes usando o Microsoft Excel. Para os dados qualitativos, usou-se o ATLAS.ti. Os resultados são discutidos na próxima seção.

7.2 RESULTADOS E DISCUSSÃO

As subseções abaixo apresentam a análise quantitativa e qualitativa do survey.

7.2.1 Análise Quantitativa

As respostas e discussões mostradas abaixo foram extraídas a partir das questões 1, 2, 3, 5, 7, 8, 9 e 10.

7.2.1.1 Perfil dos Respondentes

A maioria dos entrevistados tem menos de 40 anos de idade ($N = 99$). Este número indica uma maior adoção de SCs por jovens que cresceram com a evolução tecnológica. Os respondentes com 41 anos ou mais representavam uma pequena parte da nossa amostra ($N = 14$). Considerando a educação, quase todos os respondentes estão ou já passaram pelo ensino superior ($N = 105$). Entretanto, também houveram respondentes que estão no ensino básico ($N = 8$). Quanto ao gênero, 53,9% ($N = 61$) dos entrevistados se identificaram como homens, enquanto 46,1% ($N = 52$) se identificaram como mulheres.

Com relação à ocupação, houve a participação de vários setores da sociedade. Foram obtidas respostas de pessoas nas áreas de informática, saúde, agricultura, direito, engenharia, contabilidade, militar, marketing, educação, estética, aviação, relações internacionais e jornalismo. Este resultado demonstra que, de acordo com a expectativa, os SCs estão sendo utilizados por pessoas de diferentes áreas além da computação.

7.2.1.2 SCs ou função de voz

Com relação aos SCs utilizados pelos respondentes, é mostrado na Figura 7.1 que o sistema mais utilizado é o Google Assistant ($N = 72$). O segundo mais utilizado foi a Siri, com 58 respondentes. É importante mencionar que, nesta questão, os respondentes podiam selecionar mais de uma opção. Foi observado que estes sistemas (Google Assistant e Siri) são aplicações nativas instaladas no Android e no iOS, respectivamente. Isto pode justificar sua alta utilização. Outros SCs utilizados foram Alexa, Cortana, Google Home Devices, Amazon Echo Devices, Bixby, Kinect, Google Now, e Alice.

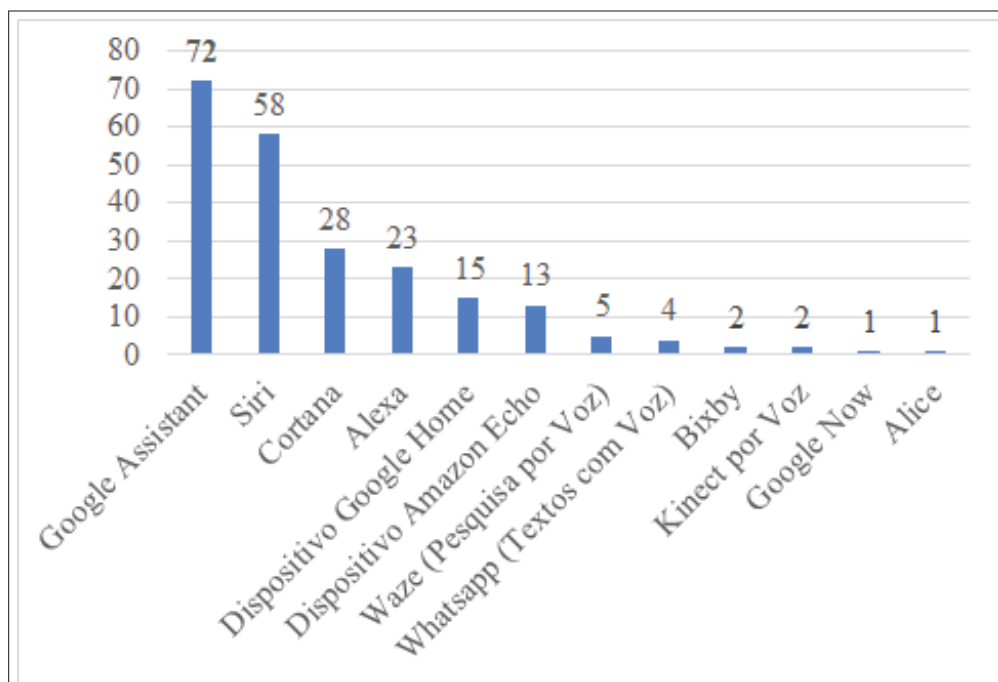


Figura 7.1: SCs usados pelos respondentes.

Também foi verificado o uso de funções de voz em aplicações que não têm voz como interação primária, como busca de voz no Waze (N = 5) e voz para escrever textos no Whatsapp (N = 4). Isto mostra a possibilidade da voz ser uma alternativa secundária às aplicações que utilizam outros tipos de interações em primeiro lugar.

7.2.1.3 Atividade realizada

De acordo com a Figura 7.2, 91 respondentes usam ou usavam os SCs para atividades diárias, como busca na Internet. O entretenimento é a segunda atividade mais realizada, seguida pela educação e aprendizagem. Estas respostas revelam o uso de SCs em diferentes contextos, tais como atividades, diversão, aprendizagem, saúde e religião.

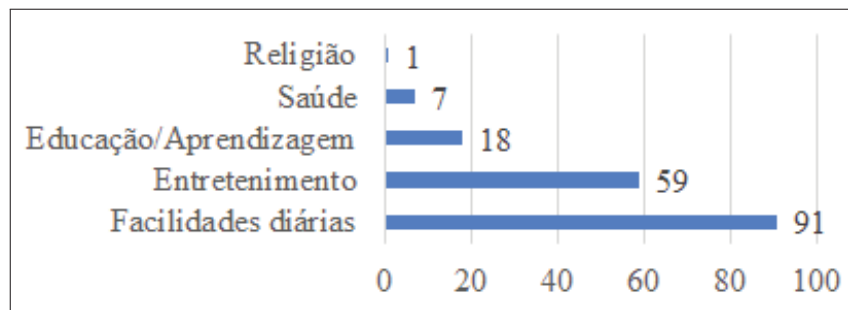


Figura 7.2: Atividades realizadas pelos respondentes.

7.2.1.4 Frequência de uso

A Figura 7.3 ilustra a frequência do uso dos respondentes. Foi observado que eles usam pouco (28,32%) ou não usam mais o SC (25,66%). Este resultado revela que os SCs apresentam dificuldades na aceitação de seus usuários. Moore et al. (2017) mostram que as falhas durante o uso de SCs contribuem para a baixa aceitação. Além disso, uma questão que poderia evidenciar este resultado é a dificuldade do usuário em se lembrar de usar a opção de voz, já que outros tipos de interação são mais convencionais, como o toque e os periféricos.

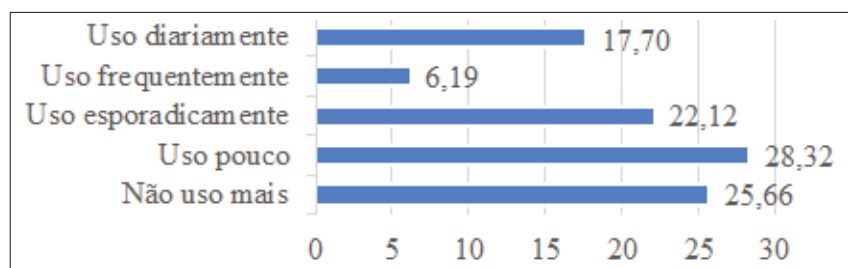


Figura 7.3: Frequência de uso dos respondentes.

7.2.1.5 *Uso geral*

Com relação à classificação de uso geral, a Figura 7.4 ilustra que a maioria dos respondentes afirmou que o uso é fácil (33,63%, N = 38) ou um pouco fácil (31,86%, N = 36). Entretanto, uma parte relevante deles classificou o uso como moderado (24,78%, N = 28), um pouco difícil (8,85%, N = 10), ou difícil (0,88%, N = 1). Embora estes respondentes sejam a minoria, é importante dar atenção a estes usuários porque o uso "difícil" pode motivar o abandono dos SCs.

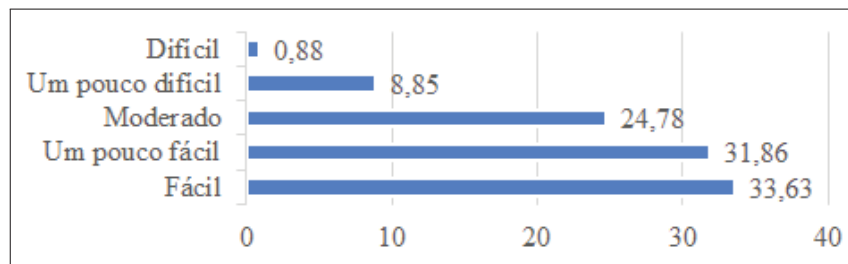


Figura 7.4: Uso geral dos respondentes.

7.2.1.6 *Realizando as tarefas com a voz*

Os resultados revelam que a maioria dos respondentes (63,72%, N = 72) precisa repetir os comandos para fazer suas ações nos SCs. Uma questão que contribui para este resultado é o reconhecimento da voz. Apesar do alto uso e investimento em SCs, algumas vezes, o sistema não reconhece comandos básicos fornecidos pelos usuários (Moore, 2017). Portanto, há necessidade de repetir o comando até que o sistema o reconheça e execute alguma ação. Além disso, 40 respondentes afirmaram que poderiam fazer o que quisessem sem ter que repetir o comando, e um respondente precisa procurar por tutoriais na Internet para usar o SC.

7.2.1.7 *Intuitividade do SC*

A Figura 7.5 ilustra os resultados em relação à intuitividade dos SCs. Uma parcela dos respondentes (35,40%, N = 40) não discorda nem concorda que os SCs são intuitivos. Somada à parcela que discorda parcialmente (12,39%, N = 14) e discorda fortemente (5,31%, N = 6), o número chega a 60 respondentes. Este resultado mostra que os SCs apresentam problemas de intuitividade, ou seja, os usuários às vezes não sabem o que fazer ou falar com o sistema. Este problema pode ocorrer porque alguns SCs não apresentam explicitamente os comandos reconhecidos, tutoriais, correção ou prevenção de erros, e os usuários se sentem perdidos.

7.2.1.8 *Preferência dos respondentes*

Os resultados revelam que 53,10% (N = 60) preferem usar outra opção em primeiro lugar do que a interação por voz. Ainda assim, quando adicionado aos respondentes que

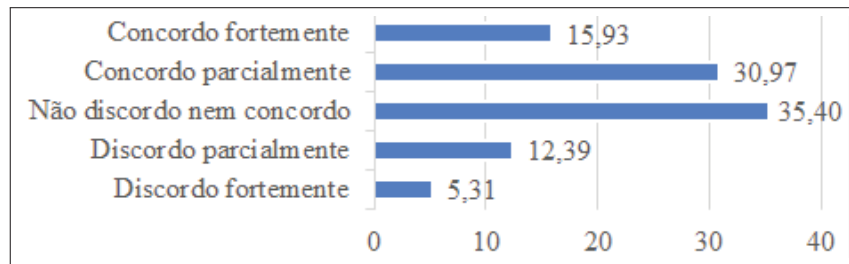


Figura 7.5: Opinião dos respondentes sobre a intuitividade dos SCs.

utilizam apenas outra opção, este número chega a 74,34% (N = 84). Os resultados evidenciam a preferência dos respondentes pelo uso de interações tradicionais em vez de interação por voz. Além disso, os resultados mostram que os SCs precisam cativar seus usuários para usar a voz. A consolidação das interações tradicionais e sua ampla utilização podem influenciar a preferência dos respondentes em optar pela interação que já estão acostumados.

7.2.1.9 Taxa de satisfação

Em relação à taxa de satisfação sobre os SCs, a Figura 7.6 ilustra os resultados obtidos. A maioria dos respondentes afirmou que o uso de SCs é mais satisfatório do que frustrante (42,48%, N = 48). Entretanto, apenas 15,04% (N = 17) responderam que o uso é fortemente satisfatório. Apesar das perguntas anteriores mostrarem alguns problemas, os respondentes acreditam que o uso de SCs, em geral, é satisfatório.

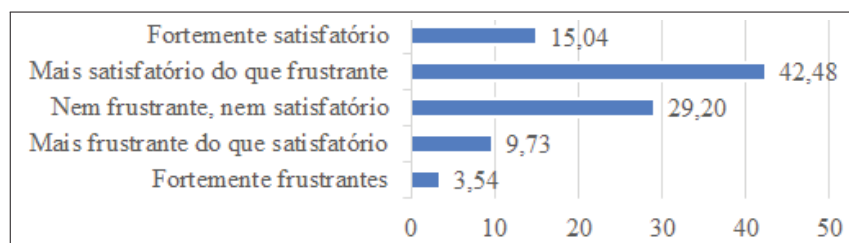


Figura 7.6: Opinião dos respondentes sobre a satisfação em relação aos SCs.

7.2.2 Análise Qualitativa

Foram apresentadas quatro perguntas abertas no survey, a 4, 6, 11, e 12. Para analisar as respostas destas perguntas, foram utilizados procedimentos do método *Grounded Theory* (GT) (Strauss e Corbin, 2014). Foram realizadas a 1ª fase (codificação aberta) e 2ª fase (codificação axial) para gerar códigos e para classificá-los. Não foi utilizada a 3ª fase (codificação seletiva) do GT porque é necessário mais estudos, opiniões e análises para gerar uma teoria sobre o tema da pesquisa.

Foram identificados alguns pontos positivos dos SCs no contexto brasileiro relacionados a benefícios ligados aos usuários. Estes benefícios demonstram que os SCs podem ajudar as pessoas que têm dificuldades para realizar suas tarefas, proporcionando acessibilidade (ver as

citações do R55 e R66 abaixo). Isto acontece porque, através da voz, os usuários cegos ou com deficiência física podem interagir com o sistema. Da mesma forma, em um contexto geral, os SCs podem proporcionar otimização de tempo, já que a voz é uma característica humana que pode ser usada com praticidade e agilidade em aplicações e sistemas (ver a citação do R78 abaixo).

- *"Há também um grande potencial para ajudar as pessoas que têm medo de usar um sistema baseado em uma interface gráfica tradicional."* (R55)
- *"Eles (SCs) permitem que pessoas com deficiências motoras possam interagir com o software de forma satisfatória."* (R66)
- *"Não tive dificuldades, mas acredito que os meios de uso atuais satisfazem minhas necessidades em relação à praticidade e ao tempo gasto na sua utilização."* (R78)

Além disso, os SCs podem fornecer outras características positivas, como o reconhecimento de voz em tempo real (ver a citação do R1 abaixo). No contexto brasileiro, também foi observado que os entrevistados percebem o investimento em SCs e que as empresas que desenvolvem estes sistemas estão aprimorando-os para fornecer novos recursos e utilidades (ver a citação do R12 abaixo).

- *"Reconhecimento 24h, então se o iPad estiver longe, basta gritar "Ei, Siri" que você descobre onde ele está e ainda consegue resolver alguma tarefa rápida".* (R1)
- *"Eles (SCs) estão sempre com novos recursos".* (R12)

Os respondentes identificaram vários recursos e utilidades de SCs que são aplicados no contexto brasileiro. Por exemplo, a automação doméstica (ver a citação do R76 abaixo), a possibilidade de usar SCs ao realizar outras atividades (ver a citação do R63 abaixo), e maior segurança no trânsito (ver a citação do R56 abaixo). Foi observado que a automação doméstica está inserida no contexto brasileiro e conta com a ajuda de SCs. Por exemplo, dispositivos físicos como o Amazon Echo e o Google Home fornecem o recurso de automação para a língua portuguesa brasileira. Além disso, os SCs ajudam quando o usuário realiza outras atividades, como lavar pratos, dirigir, ou até quando o usuário está com as mãos sujas. Com os SCs, nestas ocasiões, os usuários normalmente podem interagir, trocando músicas, fazendo ligações, enviando mensagens ou inserindo um novo endereço enquanto dirigem, por exemplo.

- *"Interação com dispositivos como TV, ar-condicionado e sensores de temperatura, portas e janelas".* (R76)
- *"Ajuda nos momentos em que nossas mãos estão ocupadas, o que, dependendo da situação, não nos permite usar o touch do celular."* (R63)
- *"Situações em que você não pode usar outras opções de interação além da voz são super positivas, por exemplo: dirigir (você não pode olhar para a tela ou digitar)." (R56)*

No entanto, foram identificados alguns problemas que os respondentes brasileiros têm quando usam SCs. Por exemplo, alguns SCs só funcionam quando conectados na Internet (ver as citações do R40 e R23 abaixo). No contexto brasileiro, a necessidade da Internet pode ser um problema para os SCs, uma vez que alguns lugares não têm acesso. Em outros casos, a conexão é lenta, colocando em risco a funcionalidade dos SCs.

- *"Quando você está offline, é difícil utilizá-lo (SC)" (R40)*
- *"Dependendo da qualidade da Internet, a interação de voz torna-se difícil." (R23)*

Além disso, os respondentes mencionaram alguns problemas de Usabilidade identificados nos SCs, tais como a dificuldade de usar o sistema inicialmente, a necessidade de outros tipos de interação para corrigir os erros de interação de voz, e a lentidão quando o usuário insere uma frase longa como um comando. Em alguns casos, os SCs não fornecem um tutorial inicial sobre como utilizá-lo. Alternativamente, quando eles fornecem, o tutorial não é totalmente adequado, o que pode gerar dificuldades iniciais em seus usuários (ver a citação do R13 abaixo). Os respondentes também relataram que os SCs não corrigem seus erros através da voz, ou seja, o usuário precisa corrigir os erros de voz com o toque, periféricos, entre outros tipos de interação (ver a citação do R109 abaixo). Ainda, os respondentes abordaram que, quando inserido uma longa entrada, o SC atrasa para fornecer uma resposta (ver a citação do R72 abaixo).

- *"Eu tive dificuldades iniciais ao usar o sistema." (R13)*
- *"Às vezes, quando o software entende mal o que eu peço, ele abre programas ou faz coisas que precisam de interação física para cancelar." (R109)*
- *"O sistema tem dificuldade de responder rapidamente quando a questão é longa." (R72)*

Os respondentes também mencionaram problemas de UX, como o usuário falar como um robô, o SC não entender o contexto da fala e a necessidade de conhecer os comandos do SC antes de utilizá-lo. Com relação à fala como um robô, os respondentes relacionaram a necessidade de falar devagar (ver a citação do R100 abaixo). Além disso, segundo os respondentes, os SCs apresentam dificuldades na interpretação de um contexto de fala, ou seja, as respostas são "prontas" e não têm uma interação real (ver a citação do R37 abaixo). Ainda, para alcançar objetivos com os SCs, os respondentes citaram a necessidade de procurar comandos e truques antes de usar (ver a citação do R15).

- *"Eu geralmente falo como um robô, muito lento." (R100)*
- *"O sistema só responde a perguntas prontas, sem interação real." (R37)*
- *"O uso é geralmente fácil (método), mas o que torna difícil é a necessidade de aprender os truques do sistema, falar de uma maneira que o robô entenda." (R15)*

Além disso, os respondentes identificaram vários problemas com reconhecimento de voz em SCs, tais como reconhecer comandos erroneamente, não reconhecer a voz do usuário ou reconhecer a voz de outras pessoas, dificuldades para usar o sistema em ambientes com ruído e dificuldades dos sistemas para reconhecer a voz quando o usuário está usando máscaras. Com relação ao reconhecimento errado de comandos, os respondentes relataram que os SCs retornam um resultado que não é esperado por eles. A razão para isso pode ser o sotaque brasileiro (ver a citação do R29 abaixo), o uso de sinônimos que não são identificados pelo SC (ver a citação do R54 abaixo), e o ambiente com ruído (ver a citação do P108 abaixo). Além disso, os respondentes detectaram falhas nos SCs quando reconhecem a voz de outras pessoas, por exemplo, quando o sistema age com um comando fornecido sem intenção por outra pessoa (ver a citação do P61 abaixo). Ainda, a pandemia causada pela COVID-19 tornou necessário o uso de máscaras pelos brasileiros, o que pode influenciar o comando de voz, já que seu uso pode abafar a fala (ver a citação do R17 abaixo).

- *"A dificuldade é que às vezes o software não entende exatamente algo dito por causa do meu sotaque."* (R29)
- *"O sistema não reconhece o comando quando eu uso sinônimos."* (R54)
- *"Às vezes, em um ambiente onde se está fazendo uma interação por voz, pode haver algum barulho, e o barulho pode atrapalhar a interação."* (R108)
- *"Quando alguém diz 'Google' perto do meu telefone, ele abre o comando de voz."* (R61)
- *"Dificuldade de entender a voz devido ao uso de máscara."* (R17)

Além disso, os respondentes mencionaram que o SC não apoia a língua portuguesa (ver a citação do R76 abaixo) e não compreende tópicos sobre as regiões brasileiras (ver a citação do R90 abaixo). Foi identificado também que os respondentes poderiam ter medo de usar SCs e ser ouvidos o tempo todo ou ter suas conversas salvas, perdendo sua privacidade (ver a citação do R66 abaixo). Ainda, os respondentes justificaram o abandono dos SCs porque não estão acostumados a usar a voz diariamente para interagir, preferindo tipos de interação convencionais (ver a citação do R78 abaixo).

- *"Dificuldade em entender o português. Por ser uma língua difícil, os assistentes em inglês sofrem um pouco."* (R76)
- *"Às vezes você tenta procurar coisas regionais brasileiras, mas o sistema não consegue entender."* (R90)
- *"Eu parei de usar depois de ver que o Google tinha salvo cada vez que eu dizia Ok Google. Achei muito surreal e bizarro; o registro sempre começava antes de eu dizer Ok Google."* (R66)
- *"Deixei de usar principalmente pelo costume com meios analógicos/touch."* (R78)

7.3 SÍNTESE DO CAPÍTULO

Neste capítulo, foram apresentados os resultados quantitativos e qualitativos de um survey realizado para identificar opiniões e possíveis dificuldades dos usuários e ex-usuário de SCs no contexto brasileiro. O survey obteve 113 respondentes de diversas áreas do conhecimento, tais como informática, saúde, direito, entre outras.

Da análise dos dados, são destacados os seguintes resultados: (i) os respondentes usaram e abandonaram as SCs devido a vários fatores, como a falta de costume no uso diário da interação por voz e a preferência pelas interações convencionais; (ii) no contexto brasileiro, os SCs apresentam vários problemas de reconhecimento de voz, como a dificuldade de reconhecer sotaques, sinônimos e tópicos regionais; (iii) a necessidade de conexão com a Internet pode ser um problema no Brasil, uma vez que alguns lugares não têm acesso ou a conexão é lenta; (iv) os SCs exigem que o usuário fale devagar, como um robô; além disso, o resultado às vezes não considera o contexto da fala, fornecendo respostas prontas; (v) os usuários têm dificuldades para aprender a usar os SCs e seus truques; às vezes, os usuários precisam pesquisar os comandos ou repetir várias vezes até atingir o objetivo; (vi) o uso de máscaras devido à pandemia da COVID-19 pode atrapalhar a interação por voz, uma vez que pode abafar a voz.

8 ESTUDO DE VIABILIDADE

Um estudo de viabilidade tem como objetivo fornecer ao pesquisador informações suficientes para continuar o trabalho (Shull et al., 2001). Este estudo teve como objetivo verificar a eficiência dos participantes ao identificar defeitos em um SC com a U2XECS, se os participantes consideravam o U2XECS fácil de usar e útil, e se pretendem usá-la no futuro.

8.1 PLANEJAMENTO

8.1.1 Contexto

Foi selecionado o aplicativo Amazon Alexa para ser avaliado pela U2XECS, visto sua popularidade e os investimentos que os produtos da Amazon estão recebendo nos últimos anos. A Amazon, com seu assistente virtual Alexa, está no topo do mercado de *smart speakers* desde 2017, e a tendência é continuar em 2021 (eMarketer, 2020). Somente na Play Store (dispositivos Android), até a data da escrita desta dissertação, o aplicativo Amazon Alexa havia sido baixado por mais de 50 milhões de usuários.

8.1.2 Seleção dos Participantes

Participaram do estudo 33 estudantes das turmas de IHC e ES dos cursos de Ciência da Computação, Informática e Sistemas de Informação da Universidade Estadual de Maringá (UEM), Universidade Estadual do Norte do Paraná (UENP) e da Universidade Estadual do Paraná (UNESPAR).

Os participantes preencheram um formulário de caracterização para categorizar seus conhecimentos como: nenhuma, baixa, média ou alta experiência em desenvolvimento de sistemas, desenvolvimento de CS e avaliação de Usabilidade/UX. Para o desenvolvimento de sistemas e desenvolvimento de CS, foi considerado:

- (a) alta experiência, participantes que tinham trabalhado em mais de quatro projetos de desenvolvimento de sistemas/SC na indústria;
- (b) experiência média, participantes que tinham trabalhado de 1 a 4 projetos de desenvolvimento de sistemas/SC na indústria;
- (c) baixa experiência, participantes que tinham trabalhado em pelo menos um projeto de desenvolvimento de sistemas/SC na indústria;
- (d) sem experiência, participantes que não tinham conhecimento prévio de sistemas/SC ou que tinham algum conhecimento, mas nenhuma experiência prática.

Para avaliação de Usabilidade/UX, foi considerado:

- (a) alta experiência, participantes que tinham trabalhado em projetos/avaliações de Usabilidade/UX na indústria;
- (b) experiência média, participantes que tinham trabalhado em projetos/avaliações de Usabilidade/UX em sala de aula;
- (c) baixa experiência, participantes que tinham conhecimento de Usabilidade/UX adquirido em palestras ou leituras;
- (d) sem experiência, participantes que não tinham conhecimento de Usabilidade/UX.

A Tabela 8.1 (segunda, terceira e quarta coluna) demonstra a categorização de cada participante. O rótulo "P" e um número identificam cada participante, por exemplo, P1 identifica o participante 1.

8.1.3 Seleção de Indicadores

Foram definidos quatro indicadores a serem avaliados neste estudo: eficiência, facilidade de uso, utilidade e intenção de uso.

- Eficiência: total de defeitos identificados pelo participante dividido pelo tempo total gasto usando U2XECS para avaliar o SC.

Após avaliar o aplicativo Amazon Alexa usando a segunda versão da U2XECS, os participantes responderam um questionário pós-estudo baseado nos indicadores do TAM (Venkatesh e Davis, 2000) sobre o uso da U2XECS. O TAM é um modelo amplamente utilizado para avaliar a aceitação de novas tecnologias (Marangunić e Granić, 2015). O TAM tem três indicadores com suas respectivas afirmativas, que são as mesmas definidas na seção 4.5.3 desta dissertação. Além das afirmativas do TAM, foi utilizada uma pergunta aberta para extrair possíveis melhorias para a U2XECS, que é: *Na sua opinião, como a U2XECS poderia ser melhorada?*

8.1.4 Instrumentação

Foram usados alguns artefatos (todos disponíveis no Google Forms) para apoiar o estudo de viabilidade: (i) formulário de consentimento (TCLE)¹, para garantir a confidencialidade e a privacidade dos dados coletados; (ii) formulário de caracterização², que tinha perguntas para caracterizar a experiência dos participantes sobre o desenvolvimento e avaliação de sistemas; (iii) folha de instruções³, contendo as tarefas a serem realizadas pelos participantes do estudo; (iv)

¹Formulário de consentimento: <https://bit.ly/3bMEHwg>

²Formulário de caracterização: <https://bit.ly/2ZvCZdT>

³Formulário de instruções: <https://bit.ly/3k9cYc8>

U2X ECS (segunda versão); e (v) questionário de pós-estudo⁴ contendo indicadores do TAM e uma pergunta aberta, para verificar o grau de concordância dos participantes e identificar melhorias para a U2X ECS.

8.1.5 Preparação

Os participantes, antes da execução, foram contextualizados sobre a avaliação de Usabilidade e UX, SCs, e a U2X ECS. Para cada turma, foi feita uma apresentação de 30 minutos sobre os tópicos mencionados. Foram fornecidos exemplos de SCs, tecnologias de avaliação de Usabilidade/UX, e o uso da U2X ECS. Ao final de cada apresentação, as dúvidas dos participantes sobre os conceitos abordados foram respondidas.

8.2 EXECUÇÃO

Devido à pandemia causada pelo COVID-19, o estudo teve que ser adaptado para o contexto online. Foi realizada uma reunião com cada turma pelo Google Meet⁵. Após a preparação dos participantes, a folha de instruções foi enviada para informar as atividades e tarefas a serem realizadas. Os participantes executaram as atividades e as tarefas e enviaram a folha de instruções contendo o nome e o tempo gasto nas tarefas. As respostas aos outros artefatos foram todas registradas pelo Google Forms. Os participantes executaram as seguintes etapas no estudo:

- PASSO 1

- Os participantes assinaram o formulário de consentimento e responderam ao questionário de caracterização;

- PASSO 2

- Pelo celular, os participantes baixaram e configuraram o aplicativo Amazon Alexa;
- Participantes anotaram o tempo inicial;
- Os participantes realizaram as seguintes tarefas no aplicativo da Amazon Alexa:
 - * Pergunte à Alexa: "Alexa, quais coisas eu posso experimentar?";
 - * Pergunte à Alexa quando é o dia da independência;
 - * Peça à Alexa para soletrar a palavra "Abóbora";
 - * Pergunte à Alexa o que significa a palavra "Medição";
 - * Pergunte à Alexa quem é "Elon Musk";
 - * Pergunte à Alexa a previsão do tempo em sua cidade;

⁴Questionário de pós-estudo: <https://bit.ly/33eyw0g>

⁵<https://meet.google.com/>

- * Pergunte à Alexa sobre o clima em Curitiba;
- * Pergunte à Alexa quantos dias faltam para o Natal;
- * Pergunte à Alexa quanto é 756 dividido por 9;
- * Pergunte à Alexa quantos reais são 25 dólares.
- Após concluir as tarefas acima, os participantes avaliaram o aplicativo Amazon Alexa através da U2XECS;
- Depois de responder a U2XECS, os participantes anotaram o tempo final;
- PASSO 3
 - Após utilizar a U2XECS, os participantes avaliaram a U2XECS quanto à sua facilidade de uso, utilidade e intenção de uso, através do questionário pós-estudo contendo indicadores TAM e a pergunta aberta;

8.3 ANÁLISE QUANTITATIVA

A Tabela 8.1 mostra os resultados gerais da avaliação dos participantes. A coluna 5 descreve o número de discrepâncias identificadas por participante. A coluna 6 mostra o número de falsos positivos. A coluna 7 mostra o número de defeitos por participante. A coluna 8 descreve o tempo gasto por cada participante para realizar as tarefas e responder a U2XECS. Finalmente, a coluna 9 mostra a relação de defeitos por hora (eficiência).

É possível verificar que os participantes identificaram entre 0 e 9 defeitos no aplicativo Amazon Alexa, gastando entre 9 e 70 minutos. P4 obteve maior eficiência em comparação com outros participantes, identificando cerca de 16,80 defeitos por hora. Foi notado que a maioria dos participantes que tiveram a maior eficiência (P2, P4, P10 e P24) tinham experiência média com avaliação de Usabilidade e UX, o que pode ter ajudado a identificar defeitos. Por outro lado, a maioria dos participantes com baixa eficiência (P3, P7, P14, P17 e P23) tinha nenhuma ou baixa experiência com avaliação de Usabilidade e UX, o que pode ter influenciado a identificação de defeitos na aplicação.

Todas as respostas dos participantes foram analisadas para considerar se eram defeitos ou não. Considerando duplicatas, os participantes descreveram 100 defeitos que identificaram no aplicativo Amazon Alexa através da U2XECS. Sem considerar as duplicatas, 48 defeitos diferentes foram identificados. Exemplos destes defeitos são "Não me senti tão confiante por causa da resposta errada dada pelo assistente", "A Alexa não retornou sugestões para corrigir a palavra que não identificou", e "Há muitos ícones que podem dificultar o aprendizado". Os comentários dos participantes que não foram considerados como defeitos foram caracterizados como falsos positivos.

Tabela 8.1: Resumo da caracterização e identificação de defeitos pelos participantes.

Part.	DS	DSC	AUUX	Disc.	Falsos Pos.	Def.	Tempo (min)	Defeito/Hora
P1	B	N	B	4	1	3	21	8,57
P2	M	N	M	10	1	9	35	15,43
P3	B	N	B	0	0	0	30	0,00
P4	N	N	N	7	0	7	25	16,80
P5	N	N	N	5	0	5	23	13,04
P6	B	N	B	4	0	4	22	10,91
P7	M	N	B	1	0	1	29	2,07
P8	M	N	B	2	0	2	52	2,31
P9	M	B	N	4	1	3	31	5,81
P10	B	N	M	6	0	6	24	15,00
P11	N	N	M	0	0	0	10	0,00
P12	N	N	N	3	0	3	67	2,69
P13	B	N	M	3	0	3	70	2,57
P14	N	N	B	2	0	2	62	1,94
P15	B	M	M	6	0	6	46	7,83
P16	N	B	B	4	0	4	22	10,91
P17	N	N	B	1	1	0	15	0,00
P18	N	N	M	6	1	4	31	7,74
P19	B	B	M	3	0	3	46	3,91
P20	B	N	B	1	0	1	23	2,61
P21	B	N	M	2	0	2	56	2,14
P22	N	N	N	1	0	1	14	4,29
P23	N	N	N	2	1	1	29	2,07
P24	B	N	M	5	1	4	15	16,00
P25	B	B	M	4	0	4	31	7,74
P26	B	N	N	4	1	3	25	7,20
P27	B	N	M	1	0	1	29	2,07
P28	B	N	M	2	0	2	9	13,33
P29	B	N	B	3	0	3	11	16,36
P30	B	N	M	2	0	2	20	6,00
P31	B	N	M	3	1	2	48	2,50
P32	B	N	M	6	1	5	46	6,52
P33	B	N	M	4	0	4	44	5,45

Legenda - **N**: Nenhuma; **B**: Baixa; **M**: Média; **DS**: Experiência em Desenvolvimento de Sistemas; **DSC**: Experiência em Desenvolvimento de Sistemas Conversacionais; **AUUX**: Experiência em Avaliação de Usabilidade/UX; **Disc**: Número de Discrepâncias; **Falsos Pos**: Número de Falsos Positivos

8.4 ANÁLISE DA ACEITAÇÃO DA U2XECS

A análise da percepção dos participantes deste estudo foi baseada nas declarações do TAM para avaliar a facilidade de uso, a utilidade percebida e as intenções de uso de nossos participantes em relação ao U2XECS. A Figura 8.1 ilustra os resultados que foram obtidos.

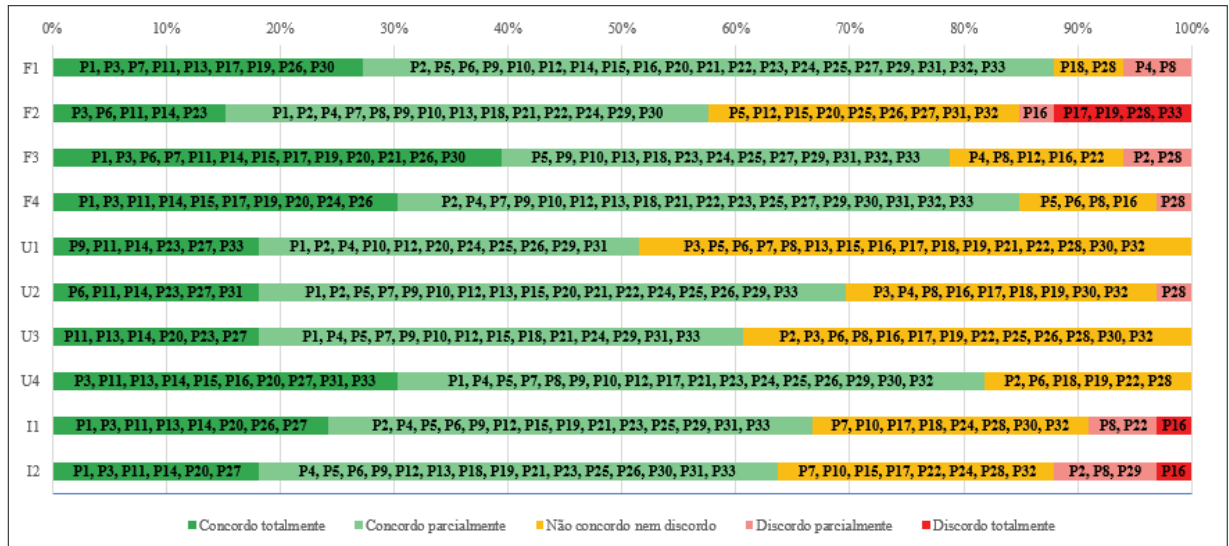


Figura 8.1: Resultados sobre a U2XECS obtidos com as declarações do TAM.

8.4.1 Facilidade de uso percebida

Os resultados indicam que os participantes acharam a U2XECS fácil de usar. Na F1, 88% dos participantes (N = 29) concordaram totalmente ou parcialmente que a interação com a U2XECS foi clara e compreensível. Na F2, 58% dos participantes (N = 19) concordaram totalmente ou parcialmente que a U2XECS não exigia muito esforço mental. Ainda, na F3, 79% dos participantes (N = 26) concordaram totalmente ou parcialmente que a U2XECS era fácil de usar. Além disso, na F4, 85% dos participantes (N = 28) concordaram totalmente ou parcialmente que a U2XECS era fácil de usar para o que eles queriam que ela fizesse, avaliar a Usabilidade e a UX de um SC.

No entanto, é possível observar que na F2, alguns participantes não deram uma resposta positiva à U2XECS. Cerca de 42% dos participantes (N = 14) não concordaram ou discordaram, ou discordaram parcialmente, ou discordaram totalmente de que a U2XECS não exigiu muito esforço mental. Este resultado indica que foi necessário um esforço mental relevante para que vários participantes respondessem a U2XECS. A quantidade de questões da segunda versão da U2XECS pode ter sido um fator que influenciou esta necessidade de esforço mental.

8.4.2 Utilidade percebida

Os resultados do TAM indicam que os participantes perceberam a utilidade da U2XECS, mas ainda não se convenceram totalmente desta utilidade. Na U1, foi identificado um equilíbrio

nas opiniões dos participantes. Cerca de 52% dos participantes (N = 17) concordaram totalmente ou parcialmente que a U2XECS poderia melhorar seu desempenho na avaliação da Usabilidade e da UX do SC. No entanto, 48% dos participantes (N = 16) não concordaram nem discordaram sobre o aumento de desempenho. Na U2, os resultados positivos estiveram mais presentes. Cerca de 70% dos participantes (N = 23) concordaram totalmente ou parcialmente que o uso da U2XECS pode aumentar sua produtividade. Na U3, é possível notar novamente um equilíbrio entre as respostas. Cerca de 61% dos participantes (N = 20) concordaram totalmente ou parcialmente que a U2XECS poderia aumentar a eficácia da avaliação da Usabilidade e da UX do SC. Por outro lado, 39% dos participantes (N = 13) não concordaram ou discordaram. Finalmente, na U4, cerca de 82% dos participantes (N = 27) consideram a U2XECS útil na avaliação da Usabilidade e da UX do SC.

As respostas negativas em relação à utilidade da U2XECS podem estar relacionadas com a experiência de nossos participantes. Como demonstrado na Tabela 8.1, alguns participantes não tinham nenhuma ou pouca experiência com avaliação de Usabilidade ou UX. Além disso, a maioria dos participantes não tinha nenhuma experiência com o desenvolvimento de SCs. Devido a estes fatores, alguns participantes podem não ter considerado a U2XECS útil.

8.4.3 Intenção de uso

Os resultados do TAM indicam que os participantes pretendem utilizar o U2XECS no futuro. No entanto, também pode ser notado que alguns participantes estavam inseguros sobre o uso futuro da U2XECS. Em I1, 67% dos participantes (N = 22) concordaram totalmente ou parcialmente que, tendo acesso à U2XECS, eles usariam a tecnologia. Por outro lado, 33% dos participantes (N = 11) não concordaram nem discordaram, discordaram parcialmente, ou discordaram totalmente com o uso futuro da tecnologia. Na I2, 64% dos participantes (N = 21) concordaram totalmente ou parcialmente que, tendo acesso à U2XECS, eles preveem que a utilizariam em outras ocasiões. Entretanto, 36% dos participantes (N = 12) não concordaram nem discordaram, ou discordaram parcialmente, ou discordaram totalmente da previsão de uso da U2XECS em outras ocasiões.

Os mesmos fatores que influenciaram o indicador de facilidade de uso e de utilidade também influenciaram as intenções de uso. A baixa experiência dos participantes, especialmente com SCs, pode ter influenciado as respostas negativas, já que alguns participantes nunca participaram de projetos de desenvolvimento de SC e não têm previsão de participar, o que influencia a opinião sobre a intenção de usar a U2XECS.

8.5 ANÁLISE QUALITATIVA

A análise qualitativa foi baseada na 1ª fase (codificação aberta) e 2ª fase (codificação axial) do GT (Strauss e Corbin, 2014). Não foi utilizada a 3ª fase (codificação seletiva) porque são necessários mais experimentos e ciclos de análise, e não era pretendido criar uma teoria. Portanto,

foram identificadas três categorias principais: benefícios do questionário, pontos negativos e sugestões de melhoria. Cada uma destas categorias está detalhada nas subseções abaixo.

8.5.1 Benefícios do questionário

Os participantes do estudo elogiaram a U2XECS e suas características, tais como a composição das afirmações, utilidade e estrutura. Os participantes gostaram da clareza das afirmações e do formato em que são exibidas (ver citação do P5). Eles também identificaram alguns pontos úteis da U2XECS e como ele pode apoiar o pesquisador que a utiliza (ver citação do P9). Além disso, a divisão estrutural entre os aspectos de Usabilidade e UX agradou os participantes (ver citação P27).

- *"Como um benefício, a U2XECS deixa claras as afirmações e o que está analisando."* [P5]
- *"A U2XECS me fez questionar sobre pontos da aplicação que às vezes ignoro sem querer."* [P9]
- *"Não tive dificuldades com a U2XECS; as afirmações foram muito claras e bem agrupadas em tópicos específicos."* [P27]

8.5.2 Pontos negativos

Os pontos negativos que os participantes identificaram na U2XECS foram sobre o tamanho do questionário e a dificuldade em entender algumas afirmações. Os participantes acharam o questionário relativamente longo (ver citação do P1), o que os deixou cansados e entediados (ver citação do P8). Além disso, alguns participantes mencionaram a dificuldade de entender algumas afirmações (ver citação do P4) e a repetição de algumas delas (ver citação do P10).

- *"O questionário é muito longo, o que o torna cansativo perto do final."* [P1]
- *"O questionário é muito grande; ele me deixou entediado."* [P8]
- *"A U2XECS é fácil de usar, mas cansativa, já que os textos são longos e às vezes difíceis de entender."* [P4]
- *"É um pouco extensa, e algumas declarações são um pouco semelhantes."* [P10]

8.5.3 Sugestões de melhoria

Os participantes contribuíram sugerindo melhorias à U2XECS em relação à funcionalidade, tipos de resposta e estrutura geral. Algumas dessas sugestões foram implementadas na terceira versão da U2XECS e são discutidas no próximo Capítulo. Outras sugestões serão melhor

analisadas para implementação futura. Os participantes sugeriram que as afirmações poderiam ser respondidas por voz (ver citação do P5). Além disso, os participantes sugeriram a redução das respostas textuais (ver citação do P32) e mencionaram que a afirmação qualitativa de cada aspecto deveria ser opcional (ver citação do P8).

- *"A U2XECS poderia implementar recursos de avaliação por voz, tais como responder a uma pergunta através da fala, o que ajudaria a encurtar o tempo de preenchimento."* [P5]
- *"Elaborar afirmações que podem substituir afirmações que requerem respostas textuais (reduzir respostas textuais), mas manter a eficácia do questionário."* [P32]
- *"A justificativa no final de cada seção do aspecto deve ser opcional."* [P8]

8.6 SÍNTESE DO CAPÍTULO

O estudo de viabilidade com a U2XECS foi realizado para verificar as opiniões dos participantes com base nos indicadores TAM - facilidade de uso, utilidade e intenções de uso - e uma pergunta para identificar melhorias. Os resultados quantitativos dos indicadores do TAM indicaram que a maioria dos participantes achou a U2XECS fácil de usar e útil, assim como possuem a intenção de usá-lo no futuro. Além disso, a partir da análise qualitativa, foi possível verificar os pontos negativos da U2XECS, principalmente em relação ao tamanho do questionário. Esses resultados serviram como base para a proposição da terceira versão da U2XECS, explicada no próximo Capítulo.

9 PROPOSTA DA TERCEIRA VERSÃO DA U2XECS

Os estudos realizados contribuíram para a identificação de melhorias em nossa tecnologia. Com base na análise do survey e no estudo de viabilidade descritos anteriormente, a U2XECS foi atualizada para sua terceira (e atual) versão. Neste capítulo, são discutidos como os resultados dos estudos influenciaram as mudanças na tecnologia.

9.1 ALTERAÇÕES

Primeiramente, baseado nos resultados do survey exploratório, foram acrescentadas duas novas afirmativas: "*O sistema pôde lidar com meu sotaque e características linguísticas*" e "*A velocidade da minha internet não influencia os resultados do sistema*".

Em relação ao estudo de viabilidade, foi verificado que as afirmativas qualitativas ao final da U2XECS podem deixar seu uso cansativo, uma vez que requer muito tempo do participante para ler todas as afirmativas e ainda tem que escrever sobre elas. Além disso, as afirmativas qualitativas ao final da U2XECS podem gerar defeitos repetidos, ou seja, defeitos já identificados ao longo do uso da U2XECS. Portanto, as afirmativas qualitativas no final da U2XECS foram removidas e deixadas apenas ao final de cada conjunto de afirmativas por aspecto de Usabilidade/UX.

Foi verificado que os pontos negativos mais significativos da tecnologia foram o tamanho do questionário e as afirmativas semelhantes. Portanto, algumas afirmativas da U2XECS foram removidas para abordar estas sugestões identificadas. O processo de retirada de afirmativas foi baseado em três escolhas: (i) afirmativas de aspectos que identificavam falsos positivos; (ii) afirmativas que não contribuíam diretamente para a identificação de defeitos; e (iii) afirmativas que poderiam ser similares.

Foi observado que as afirmativas de motivação e encantamento contribuíram para a identificação de falsos positivos. Os falsos positivos identificados com a U2XECS foram as considerações que os participantes fizeram, mas não caracterizam de fato um defeito. Por exemplo, em relação ao aspecto de encantamento, um participante respondeu: "Eu ficaria mais encantado se eu tivesse uma transcrição do discurso em tempo real". Esta discrepância não é um defeito porque a falta de transcrição em tempo real não é um aspecto negativo do sistema. Portanto, ela foi caracterizada como um falso positivo. Foram retiradas algumas afirmativas sobre os aspectos de motivação e encantamento, que são: "*Eu acho que usaria a interação por voz regularmente nesse sistema ao invés de outro tipo de interação*" e "*As tarefas que realizei no sistema com a voz fez com que eu ficasse encantado com este tipo de interação*". Analisando estas afirmativas, pode-se perceber que elas não encorajam os participantes a julgar o SC porque dependem da preferência do usuário em usar a voz para realizar as ações.

Seguindo esta lógica, todas as afirmativas restantes na U2XECS foram relidas para reduzir o tamanho do questionário. Então, foi identificada outra afirmativa para ser removida, desta vez no aspecto UX Genérica: "*Nunca tive uma experiência de interação por voz como tive ao realizar estas tarefas no sistema*". Esta pergunta foi retirada porque não julga o SC que está sendo avaliado no estudo, mas compara o uso com a experiência que o participante já tinha no passado.

Também foram analisadas todas as afirmativas que poderiam ser semelhantes. Foi excluída a afirmativa do aspecto de prazer/diversão: "*Foi muito entediante utilizar a voz para realizar essas tarefas no sistema*". Esta pergunta pode ser considerada o oposto da pergunta sobre prazer: "*Foi muito prazeroso utilizar a voz para realizar essas tarefas no sistema*". Em outras palavras, o participante que achou a interação entediante, por lógica, não a achou prazerosa.

Também foram identificadas afirmativas que poderiam gerar ambiguidade nas opiniões dos participantes. Uma delas era do aspecto de eficiência: "*Durante a minha interação por voz com o sistema tive que repetir diversos comandos*". A outra, pelo aspecto de eficácia: "*O sistema conversacional me obrigou a usar palavras-chave*". Estas afirmativas foram removidas porque os SCs, por natureza, usam palavras-chave para ser acionado (por exemplo, "Hey, Siri!" e "Ok, Google"). Portanto, isto não é um defeito deste tipo de sistema, mas sim uma característica.

Além disso, foram retiradas todos os tipos de negação de dentro das afirmativas, deixando apenas afirmativas no sentido positivo. Resumindo, as alterações para a terceira versão da U2XECS foram:

- Adição de duas novas afirmativas com base nos resultados do survey.
- Remoção das afirmativas qualitativas no final da U2XECS, deixando apenas no final de cada aspecto;
- Remoção de duas afirmativas que identificaram falsos positivos;
- Remoção de uma afirmativa que não julgou o SC em si, mas sim as experiências passadas;
- Remoção de uma afirmativa que poderia estar repetindo o significado de outra;
- Remoção de duas afirmativas que poderiam gerar opiniões ambíguas nos participantes;
- Retirada do intercalamento entre afirmativas positivas e negativas, deixando apenas afirmativas positivas.

9.2 TEMPLATE DO QUESTIONÁRIO DA TERCEIRA VERSÃO DA U2XECS

QUESTIONÁRIO DA U2XECS
Nome:
<p>– Por favor, preencha o questionário abaixo para avaliar a sua experiência na interação com o sistema conversacional. Sua resposta é importante e nos ajudará a propor melhorias por meio desta avaliação.</p> <p>– Nas questões 1 a 30, assinale apenas uma resposta por questão. Nas questões marcadas com "**", escreva o quanto desejar. Se necessário, utilize o verso dessas folhas.</p> <p>– Não há respostas certas ou erradas. O importante para nós é a sua opinião.</p> <p>– Responda as questões de múltipla escolha utilizando a seguinte escala:</p> <p>1 = discordo totalmente;</p> <p>2 = discordo parcialmente;</p> <p>3 = não discordo nem concordo;</p> <p>4 = concordo parcialmente;</p> <p>5 = concordo totalmente.</p>

Satisfação do Usuário					
	1	2	3	4	5
1. Foi fácil utilizar a voz para realizar as tarefas nesse sistema.					
2. Não precisei aprender muito sobre o sistema antes de realizar essas tarefas com a voz.					
3. Consegui me familiarizar com o sistema conversacional quando o usei pela primeira vez.					
4. Eu me senti satisfeito ao usar a voz para realizar essas tarefas.					
5. O sistema se comportou da forma como eu esperei durante a interação por voz.					
6. Eu achei fácil entender como interagir por voz no sistema.					
7. Foi fácil me tornar habilidoso ao utilizar o sistema conversacional.					
* Descreva os problemas de satisfação que você identificou no sistema.					
Eficiência					
	1	2	3	4	5
8. O sistema respondeu de maneira imediata a minha interação por voz.					
9. Eu usaria a voz com frequência para realizar essas tarefas no sistema.					
10. O ritmo da interação por voz com o sistema foi apropriado.					

11. Consegui completar minha tarefa com a interação por voz em um tempo que acredito ser razoável.					
* Descreva os problemas de eficiência que você identificou no sistema.					
Eficácia					
	1	2	3	4	5
12. O sistema reconheceu o que eu falei durante a minha interação por voz.					
13. O sistema conversacional foi capaz de se recuperar facilmente de algum erro ou equívoco que eu cometi.					
14. Eu completei minhas tarefas com interação por voz no sistema sem cometer erros					
15. Quando a entrada de dados no sistema conversacional foi inconsistente ou ambígua, o sistema solicitava mais informações.					
16. Do ponto de vista dos usuários novatos, o sistema conduziu, de uma boa maneira, a interação por voz.					
17. Do ponto de vista dos usuários avançados, o sistema permitiu uma grande quantidade de dados de entrada de uma só vez.					
18. A velocidade da minha internet não influenciou os resultados do sistema.					
* Descreva os problemas de eficácia que você identificou no sistema.					
UX Genérica					
	1	2	3	4	5
19. Realizar essas tarefas com a voz no sistema foi uma experiência boa.					
20. O sistema pôde lidar com meu sotaque e características linguísticas.					
* Descreva os problemas de UX que você identificou no sistema.					
Afeto/Emoção					
	1	2	3	4	
21. Me senti animado ao utilizar a voz para realizar essas tarefas no sistema.					
22. Me senti confiante ao utilizar a voz para interagir com este sistema.					
23. O sistema respondeu a minha interação por voz de uma maneira amigável.					
* Descreva os problemas de afeto e emoção que você identificou no sistema.					
Prazer/Diversão					
	1	2	3	4	5
24. Foi muito prazeroso utilizar a voz para realizar essas tarefas no sistema.					
25. Eu me diverti ao utilizar a voz para realizar essas tarefas no sistema.					
* Descreva os problemas de prazer e diversão que você identificou no sistema.					
Estética/Atração					

	1	2	3	4	5
26. O sistema possuía um design inovador que facilitou a realização das tarefas por meio da voz.					
27. Utilizar a voz para realizar as tarefas me atraiu para usar o sistema.					
* Descreva os problemas de estética e atração que você identificou no sistema.					
Engajamento/Fluxo					
	1	2	3	4	5
28. Utilizar a voz fez com que eu me dedicasse mais para realizar as tarefas no sistema.					
29. Eu me senti no controle do sistema durante a interação de voz.					
* Descreva os problemas de engajamento e fluxo que você identificou no sistema.					
Motivação					
	1	2	3	4	5
30. Me senti motivado ao utilizar a voz para realizar essas tarefas no sistema.					
* Descreva os problemas de motivação que você identificou no sistema.					

10 CONSIDERAÇÕES FINAIS

Esta dissertação de mestrado teve o objetivo de apresentar a U2XECS (*Usability and User eXperience Evaluation of Conversational Systems*) como apoio a pesquisadores e desenvolvedores no processo de avaliação de Sistemas Conversacionais. Desse modo, por meio de um questionário, a U2XECS fornece uma série de afirmativas que abordam diferentes aspectos de Usabilidade e UX: satisfação de usuário, eficiência, eficácia, UX genérica, afeto/emoção, prazer/diversão, estética/atração, engajamento/fluxo e motivação. Além disso, foram realizados dois MSLs e três estudos experimentais para identificar resultados que poderiam contribuir com a evolução da tecnologia.

Inicialmente, foi realizado um MSL para identificar as tecnologias de avaliação de Usabilidade e/ou UX usadas em NUIs. Os resultados mostraram que a maioria das tecnologias foca em apenas um critério, Usabilidade ou UX. Além disso, independentemente de a Usabilidade ser avaliada em conjunto com UX ou não, a maioria das avaliações identificadas não são específicas para o contexto NUI. Em outras palavras, elas podem ser usadas para avaliar qualquer sistema ou serviço. Quando são específicas, as tecnologias são projetadas apenas para o trabalho em questão, sem fornecer um padrão de replicabilidade. Além disso, a maior parte dos estudos encontrados focaram seus resultados na análise quantitativa. A partir dos resultados deste MSL abrangente, foi identificado o gap e a necessidade de propor uma tecnologia para um tipo de NUI específica, a voz. Com os resultados do primeiro MSL, foi proposta a primeira versão da U2XECS. Essa versão foi submetida a um estudo exploratório com dois participantes. Os resultados mostraram que a U2XECS obteve uma boa aceitação, e também identificaram algumas orientações fornecidas para melhorar a tecnologia. A principal orientação foi a necessidade de inserir mais afirmativas, o que motivou a realização do segundo MSL, desta vez específico para interações por voz e Sistemas Conversacionais.

Os resultados do segundo MSL mostraram que os pesquisadores estão criando os seus próprios questionários de avaliação, específicos para o seu estudo, e que não passam por nenhum processo de validação ou avaliação empírica. Além disso, foi notado também uma quantidade de avaliações apenas de Usabilidade, enquanto a UX ainda não possui aspectos bem definidos neste contexto. Foi identificada também a utilização de tecnologias genéricas, as quais não levam em consideração as especificidades que uma interação por voz possui. Ainda, a maioria das tecnologias realizam apenas análise quantitativa, que não levam em consideração possíveis pensamentos subjetivos dos usuários que podem contribuir com a qualidade do software. A partir da análise das tecnologias retornadas no MSL, foi proposta a segunda versão da U2XECS.

Após a proposição da segunda versão, foi realizado um survey exploratório, que teve como objetivo verificar o uso e as dificuldades de usuários e ex-usuários de Sistemas Conversacionais. Os resultados mostraram que os entrevistados usam e abandonam os SCs

devido a vários fatores, tais como a falta de costume no uso diário da interação por voz e a preferência pelas interações convencionais. Além disso, foi verificado também diversos relatos relacionados a problemas de reconhecimento de voz, tais como dificuldade para reconhecer sotaques, sinônimos e tópicos regionais. A necessidade de SCs de conexão com a Internet também foi verificada como um fator problemático, uma vez que alguns lugares não têm acesso à Internet ou a conexão é lenta. Os resultados do survey auxiliaram na criação de mais duas afirmativas para a U2XECS, relacionadas ao sotaque e à conexão com a internet.

Após a realização do survey, foi feito um estudo de viabilidade com a U2XECS. O estudo foi realizado com estudantes, e teve como objetivo utilizar a U2XECS para avaliar o aplicativo Amazon Alexa. Após a avaliação, os estudantes responderam o TAM sobre a U2XECS e uma questão aberta sobre possíveis melhorias. Os resultados quantitativos dos indicadores do TAM indicaram que a maioria dos participantes achou a U2XECS fácil de usar e útil, e possuem a intenção de usá-la no futuro. Além disso, a partir da análise qualitativa, foi verificado os pontos negativos da U2XECS, principalmente em relação ao tamanho do questionário. Portanto, a partir dos resultados do survey exploratório e do estudo de viabilidade, foi proposta a terceira (e atual) versão da U2XECS, uma versão menor e mais otimizada, sem perder a especificidade para Sistemas Conversacionais.

10.1 CONTRIBUIÇÕES

As principais contribuições desta dissertação de mestrado são apresentadas a seguir:

- Dois MSLs que identificaram diferentes gaps de pesquisa, tanto no contexto abrangente de Interface Natural de Usuário, como o de interação baseada em voz;
- Criação da tecnologia U2XECS, que auxilia pesquisadores e desenvolvedores na avaliação conjunta de Usabilidade e UX de Sistemas Conversacionais;
- Divulgação científica, compartilhando os dados de estudos experimentais e o processo evolutivo para se chegar na versão atual da U2XECS.

10.2 PERSPECTIVAS PARA TRABALHOS FUTUROS

A seguir, são apresentadas algumas perspectivas futuras da pesquisa a serem desenvolvidas:

- **Atualização do segundo MSL:** realização de uma atualização do MSL específico para sistemas de interação por voz, abordando tanto a técnica de pesquisa em base de dados, como a de *snowballing*;
- **Criação da ferramenta de análise:** desenvolvimento de uma ferramenta que possa analisar os dados específicos de uma avaliação que utiliza a U2XECS, auxiliando mais ainda o trabalho de pesquisadores e desenvolvedores.

- **Avaliação e evolução da U2XECS:** realização de estudos experimentais para identificar melhorias e propor novas versões para a tecnologia. Além disso, planeja-se também inserir aspectos de acessibilidade na U2XECS.

10.3 PUBLICAÇÕES

GUERINO, G. C.; VALENTIM, N. M. C. **'Is anybody there?': Exploring the use and difficulties of Brazilians with Conversational Systems.** In: Brazilian Symposium on Human Factors in Computing Systems, 2020, Diamantina, Minas Gerais. Proceeding of the XIX Brazilian Symposium on Human Factors in Computing Systems, 2020. p. 1-6. DOI: 10.1145/3424953.3426649

GUERINO, G. C.; VALENTIM, N. M. C. **Usability and User eXperience Evaluation of Conversational Systems: A Systematic Mapping Study.** In: Brazilian Symposium on Software Engineering, 2020, Natal, Rio Grande do Norte. Proceedings of the XXXIV Brazilian Symposium on Software Engineering, 2020. p. 1-10. DOI: 10.1145/3422392.3422421

GUERINO, G. C.; VALENTIM, N. M. C. **Usability and User Experience Evaluation of Natural User Interfaces: A Systematic Mapping Study.** IET Software (Online), v. 14, n. 5, pp. 451-467, 2020. DOI: 10.1049/iet-sen.2020.0051

GUERINO, G. C.; VALENTIM, N. M. C. **Evaluating a Voice-Based Interaction: A Qualitative Analysis.** In: Brazilian Symposium on Human Factors in Computing Systems, 2019, Vitória. Proceedings of the XVIII Brazilian Symposium on Human Factors in Computing Systems, 2019. pp. 1-4. DOI: 10.1145/3357155.3360472

REFERÊNCIAS

- Almahri, F. A. J., Bell, D. e Merhi, M. (2020). Understanding student acceptance and use of chatbots in the united kingdom universities: A structural equation modelling approach. Em *2020 6th International Conference on Information Management, ICIM '20*, páginas 284–288. IEEE.
- Amith, M., Zhu, A., Cunningham, R., Lin, R., Savas, L., Shay, L., Chen, Y., Gong, Y., Boom, J., Roberts, K. e Tao, C. (2019). Early usability assessment of a conversational agent for hpv vaccination. *Studies in Health Technology and Informatics*, 257:17–23.
- Ashok, V., Puzis, Y., Borodin, Y. e Ramakrishnan, I. (2017). Web Screen Reading Automation Assistance Using Semantic Abstraction. Em *Proceedings of the 22Nd International Conference on Intelligent User Interfaces*, páginas 407–418, Limassol, Chipre.
- Azeta, A. A., Inam, I. A. e Daramola, O. (2018). A voice-based e-examination framework for visually impaired students in open and distance learning. *Turkish Online Journal of Distance Education*, 19(2):34–46.
- Bargas-Ávila, J. A. e Hornbæk, K. (2011). Old wine in new bottles or novel challenges: A critical analysis of empirical studies of user experience. Em *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '11*, página 2689–2698, New York, NY, USA. Association for Computing Machinery.
- Basili, V. R. e Rombach, H. D. (1988). Towards a Comprehensive Framework for Reuse: A Reuse-Enabling Software Evolution Environment. Relatório Técnico 90-47, University of Maryland, Maryland.
- Bernhaupt, R., Palanque, P., Winckler, M. e Navarre, D. (2007). Usability study of multi-modal interfaces using eye-tracking. Em Baranauskas, C., Palanque, P., Abascal, J. e Barbosa, S. D. J., editores, *Human-Computer Interaction – INTERACT 2007*, páginas 412–424, Berlin, Heidelberg. Springer Berlin Heidelberg.
- Bhuiyan, M. e Picking, R. (2011). A gesture controlled user interface for inclusive design and evaluative study of its usability. *Journal of Software Engineering and Applications*, 4:513–521.
- Bickmore, T. W., Mitchell, S. E., Jack, B. W., Paasche-Orlow, M. K. e amd Julie O'Donnell, L. M. P. (2010). Response to a relational agent by hospital patients with depressive symptoms. *Interacting with Computers*, 22:289 – 298.
- Blake, J. (2011). *Natural User Interfaces in .NET*. Manning Publications.

- Brooke, J. (1996). Sus: A “quick and dirty” usability scale. Em JORDAN, P. W. e *et al.*, editores, *Usability evaluation in industry*, páginas 1–7, London. Taylor&Francis.
- Bødker, S. (2006). When Second Wave HCI Meets Third Wave Challenges. Em *Proceedings of the 4th Nordic Conference on Human-computer Interaction: Changing Roles*, Oslo, Noruega.
- Caggianese, G., Gallo, L. e Pietro, G. D. (2014). Design and Preliminary Evaluation of a Touchless Interface for Manipulating Virtual Heritage Artefacts. Em *Proceedings of 2014 Tenth International Conference on Signal-Image Technology and Internet-Based Systems*, páginas 493–500, Marrakech, Marrocos.
- Candello, H., Pinhanez, C., Pichiliani, M., Cavalin, P., Figueiredo, F., Vasconcelos, M. e Do Carmo, H. (2019). The effect of audiences on the user experience with conversational interfaces in physical spaces. Em *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, página 1–13, New York, NY, USA. ACM.
- Chang, J. C., Lien, A., Lathrop, B. e Hees, H. (2009). Usability evaluation of a volkswagen group in-vehicle speech system. Em *Proceedings of the 1st International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, AutomotiveUI '09, páginas 137—144, New York, NY, USA. ACM.
- Chatzidaki, E. e Xenos, M. (2017). A Case Study on Learning Through Natural Ways of Interaction. Em *Proceedings of 2017 IEEE Global Engineering Education Conference (EDUCON)*, páginas 746–753, Atenas, Grécia.
- Cohen, M. H., Giangola, J. P. e Balogh, J. (2004). *Voice User Interface Design*. Addison-Wesley.
- Cowan, B. R., Pantidi, N., Coyle, D., Morrissey, K., Clarke, P., Al-Shehri, S., Earley, D. e Bandeira, N. (2017). “what can i help you with?”: Infrequent users’ experiences of intelligent personal assistants. *MobileHCI '17*, página 1–12, New York, NY, USA. ACM.
- Deng, S., Jiang, N., Chang, J., Guo, S. e Zhang, J. J. (2017). Understanding the Impact of Multimodal Interaction using Gaze Informed Mid-air Gesture Control in 3D Virtual Objects Manipulation. *International Journal of Human-Computer Studies*, 105:68–80.
- Di Nuovo, A., Broz, F., Wang, N., Belpaeme, T., Cangelosi, A., Jones, R., Esposito, R., Cavallo, F. e Dario, P. (2018). The multi-modal interface of robot-era multi-robot services tailored for the elderly. *Intelligent Service Robotics*, 11(1):109–126.
- d’Ornellas, M. C., Cargnin, D. J. e Prado, A. L. C. (2015). Evaluating the Impact of Player Experience in the Design of a Serious Game for Upper Extremity Stroke Rehabilitation. *Studies Health Technology Information*, 216:363–367.

- Dutton, R. T., Foster, J. C., Jack, M. A. e Stentiford, F. W. (1993). Identifying usability attributes of automated telephone services. Em *Third European Conference on Speech Communication and Technology*, EUROSPEECH '93. ISCA Archive.
- Eckert, M., Gómez-Martinho, I., Meneses, J. e Martínez, J. F. (2017). New Approaches to Exciting Exergame-Experiences for People with Motor Function Impairments. *Sensors*, 17(2):1–22.
- Economou, D., Doumanis, I., Argyriou, L. e Georgalas, N. (2017). User Experience Evaluation of Human Representation in Collaborative Virtual Environments. *Personal and Ubiquitous Computing*, 21(6):989–1001.
- eMarketer (2020). Amazon maintains convincing lead in us smart speaker market.
- Falcão, C., Lemos, A. C. e Soares, M. (2015). Evaluation of Natural User Interface: A Usability Study Based on the Leap Motion Device. *Procedia Manufacturing*, 3:5490–5495.
- Farinazzo, V., Salvador, M., Kawamoto, A. L. S. e de O. Neto, J. S. (2010). An empirical approach for the evaluation of voice user interfaces. Em Matrai, R., editor, *User Interfaces*. IntechOpen, Rijeka.
- Fernandez Martinez, F., Blazquez, J., Ferreiros, J., Barra, R., Macias-Guarasa, J. e Lucas-Cuesta, J. M. (2008). Evaluation of a spoken dialogue system for controlling a hifi audio system. Em *2008 IEEE Spoken Language Technology Workshop*, páginas 137–140, USA. IEEE Computer Society.
- Fuchsberger, V., Moser, C. e Tscheligi, M. (2012). Values in Action (ViA): Combining Usability, User Experience and User Acceptance. Em *CHI '12 Extended Abstracts on Human Factors in Computing Systems*, Austin, Texas, EUA.
- Grand View Research, Inc (2020). Intelligent virtual assistant market size, share & trends analysis report by product (chatbot, smart speakers), by technology, by application (bfsi, healthcare, education), by region, and segment forecasts, 2020 - 2027. <https://www.grandviewresearch.com/press-release/global-intelligent-virtual-assistant-industry>.
- Guimarães, M. D. P., Martins, V. F. e Brega, J. R. F. (2012). A Software Development Process Model for Gesture-based Interface. Em *Proceedings of IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, páginas 2985–2990, Seul, Coréia do Sul.
- Gürkök, H., Hakvoort, G., Poel, M. e Nijholt, A. (2011). User expectations and experiences of a speech and thought controlled computer game. Em *Proceedings of the 8th International Conference on Advances in Computer Entertainment Technology*, ACE '11, páginas 1–6, New York, NY, USA. ACM.

- Gürkök, H., Hakvoort, G., Poel, M. e Nijholt, A. (2017). Meeting the expectations from brain-computer interfaces. *Computers in Entertainment*, 15(3):1–10.
- Gustavsson, P., Syberfeldt, A., Brewster, R. e Wang, L. (2017). Human-robot Collaboration Demonstrator Combining Speech Recognition and Haptic Control. *Procedia CIRP*, 63:396–401.
- Hara, S., Kitaoka, N. e Takeda, K. (2010a). Automatic detection of task-incompleted dialog for spoken dialog system based on dialog act n-gram. Em *Proceedings of the 11th Annual Conference of the International Speech Communication Association*, INTERSPEECH'10, páginas 3034–3037.
- Hara, S., Kitaoka, N. e Takeda, K. (2010b). Estimation method of user satisfaction using n-gram-based dialog history model for spoken dialog system. Em *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*, páginas 78–83, Valletta, Malta. European Language Resources Association (ELRA).
- Hassenzahl, M. (2004). The Interplay of Beauty, Goodness, and Usability in Interactive Products. *Human-Computer Interaction*, 19:319–349.
- Hassenzahl, M., Burmester, M. e Koller, F. (2003). Attrakdiff: Ein fragebogen zur messung wahrgenommener hedonischer und pragmatischer qualität. Em *Mensch & Computer*, páginas 187–196. Vieweg+Teubner Verlag.
- Hassenzahl, M., Schobel, M. e Trautmann, T. (2008). How Motivational Orientation Influences the Evaluation and Choice of Hedonic and Pragmatic Interactive Products: The Role of Regulatory Focus. *Interacting with Computers*, 20(4-5):473–479.
- Hassenzahl, M. e Tractinsky, N. (2006). User Experience - A Research Agenda. *Behaviour & Information Technology*, 25(2):91–97.
- Hone, K. e Graham, R. (2000). Towards a tool for the subjective assessment of speech system interfaces (sassi). *Natural Language Engineering*, 6(3-4):287–303.
- Huerta, O., Sánchez, J. A., Fuentes, S. e Cervantes, O. (2011). Speak up your mind: Using speech to capture innovative ideas on interactive surfaces. Em *Proceedings of the 10th Brazilian Symposium on Human Factors in Computing Systems and the 5th Latin American Conference on Human-Computer Interaction*, IHC+CLIHC '11, páginas 202—211, Porto Alegre, Brazil. Brazilian Computer Society.
- Iannella, R. (2001). Graphical user interface evaluation for messaging and directory systems.
- Ijsselstein, W. A., de Kort, Y. A. W. e Poels, K. (2013). *The Game Experience Questionnaire*. Technische Universiteit Eindhoven.

- ISO 9241-11 (1998). *Ergonomic requirements for office work with visual display terminals (VDTs) - Part 11: Guidance on usability*. International Organization for Standardization.
- ISO 9241-210 (2019). *Ergonomics of Human System Interaction - Part 210: Human-Centered Design for Interactive Systems*. International Organization for Standardization.
- ISO/IEC 25010 (2011). *Systems and software engineering - SquaRE - Software product Quality Requirements and Evaluation: System and Software Quality Models*. International Organization for Standardization.
- Ivory, M. Y. e Hearst, M. A. (2001). The State of the Art in Automating Usability Evaluation of User Interfaces. *ACM Computing Surveys (CSUR)*, 33(4):470–516.
- Jackson, S. A. e Marsh, H. W. (1996). Development and Validation of a Scale to Measure Optimal Experience: The Flow State Scale. *Journal of Sport and Exercise Psychology*, 18(1):17–35.
- Jeong, Y., Lee, J. e Kang, Y. (2019). Exploring effects of conversational fillers on user perception of conversational agents. Em *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI EA '19, página 1–6, New York, NY, USA. Association for Computing Machinery.
- Jordan, P. W. (1998). *An Introduction to Usability*. Londres: Taylor & Francis.
- Kafle, S. e Huenerfauth, M. (2017). Evaluating the usability of automatically generated captions for people who are deaf or hard of hearing. Em *Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility*, ASSETS '17, página 165–174, New York, NY, USA. ACM.
- Kalckert, A. e Ehrsson, H. H. (2014). The moving rubber hand illusion revisited: Comparing movements and visuotactile stimulation to induce illusory ownership. *Consciousness and Cognition*, 26:117 – 132.
- Kamitis (2016). Intelligent personal assistant: Products, technologies and market. [Online; accessed 15/06/2020].
- Kazuma, T., Yoshida, E., Yu, Y. e He, A. (2016). PseudoHandWriting: New Approach for Oral Presentation to have Both Advantages of Slide and Handwriting. Em *Proceedings of the 30th International Conference on Advanced Information Networking and Applications Workshops (WAINA)*, páginas 461–465, Crans-Montana.
- Kirakowski, J. e Corbett, M. (1993). SUMI: The Software Usability Measurement Inventory. *British Journal of Educational Technology*, 24(3):210–212.

- Kitchenham, B. e Charters, S. (2007). Guidelines for Performing Systematic Literature Reviews in Software Engineering. Relatório Técnico EBSE-2007-01, Software Engineering Group Department of Computer Science Keele University, Keele, Reino Unido.
- Kocaballi, A. B., Coiera, E. e Berkovsky, S. (2020). Revisiting habitability in conversational systems. Em *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI EA '20, página 1–8, New York, NY, USA. ACM.
- Kocaballi, A. B., Laranjo, L. e Coiera, E. (2018). Measuring User Experience in Conversational Interfaces: A Comparison of Six Questionnaires. Em *Proceedings of the 32nd International British Computer Society Human Computer Interaction Conference*, páginas 1–12, Belfast, Northern Ireland. BCS Learning & Development Ltd.
- Kusumaningayu, F. e Ayu, M. A. (2017). A web accessing tool for blind and visually impaired people using bahasa indonesia. Em *2017 Second International Conference on Informatics and Computing (ICIC)*, páginas 1–6, USA. IEEE Computer Society.
- Laranjo, L., Dunn, A. G., Tong, H. L., Kocaballi, A. B., Chen, J., Bashir, R., Surian, D., Gallego, B., Magrabi, F., Lau, A. Y. S. e Coiera, E. (2018). Conversational agents in healthcare: a systematic review. *Journal of the American Medical Informatics Association*, 25(9):1248–1258.
- Laugwitz, B., Held, T. e Schrepp, M. (2008). Construction and Evaluation of a User Experience Questionnaire. Em Holzinger, A., editor, *HCI and Usability for Education and Work*, páginas 63–76, Berlim.
- Lavie, T. e Tractinsky, N. (2004). Assessing Dimensions of Perceived Visual Aesthetics of Web Sites. *International Journal of Human-Computer Studies*, 60(3):269–298.
- Lazar, J., Feng, J. H. e Hochheiser, H. (2017). *Research Methods in Human-Computer Interaction*. Morgan Kaufmann Publishers.
- Leap Motion (2013). Leap motion. <https://www.leapmotion.com>. Acessado em 28/03/2019.
- Lee, J., Lee, C. e Kim, G. J. (2017). Vouch: Multimodal touch-and-voice input for smart watches under difficult operating conditions. *Journal on Multimodal User Interfaces*, 11(3):289–299.
- Li, A. X., Lou, X., Hansen, P. e Peng, R. (2016). On the Influence of Distance in the Interaction With Large Displays. *Journal of Display Technology*, 12(8):840–850.
- Lovato, S. e Piper, A. M. (2015). “siri, is this you?”: Understanding young children’s interactions with voice input systems. Em *Proceedings of the 14th International Conference on Interaction Design and Children*, IDC '15, página 335–338, New York, NY, USA. ACM.

- Macaranas, A., Antle, A. N. e Riecke, B. E. (2015). What is Intuitive Interaction? Balancing Users' Performance and Satisfaction with Natural User Interfaces. *Interacting with Computers*, 27(3):357–370.
- Madan, A. e Dubey, S. K. (2012). Usability Evaluation Methods: A Literature Review. *International Journal of Engineering Science and Technology (IJEST)*, 4(2):590–599.
- Mafrá, S. N. e Travassos, G. H. (2006). Estudos Primários e Secundários Apoiando a Busca por Evidência em Engenharia de Software. Relatório Técnico ES 687/06, Universidade Federal do Rio de Janeiro, Rio de Janeiro.
- Marangunić, N. e Granić, A. (2015). Technology acceptance model: a literature review from 1986 to 2013. *Universal Access in the Information Society*, 14:81–95.
- McCarthy, J., Wright, P., Wallace, J. e Dearden, A. (2006). The Experience of Enchantment in Human–Computer Interaction. *Personal and Ubiquitous Computing*, 10(6):369–378.
- McTear, M., Callejas, Z. e Griol Barres, D. (2016). *The Conversational Interface: Talking to Smart Devices*. Springer.
- Mehrabian, A. e Russel, J. A. (1974). An Approach to Environmental Psychology. *The MIT Press*.
- Microsoft (2010). Kinect for windows. <https://developer.microsoft.com/pt-br/windows/kinect>. Acessado em 21/05/2019.
- Microsoft (2014). Kinect for windows sdk 2.0. <https://www.microsoft.com/en-us/download/details.aspx?id=44561>. Acessado em 28/03/2019.
- Miguel-Hurtado, O., Blanco-Gonzalo, R., Guest, R. e Lunerti, C. (2016). Interaction evaluation of a mobile voice authentication system. Em *2016 IEEE International Carnahan Conference on Security Technology (ICCST)*, páginas 1–8, USA. IEEE Computer Society.
- Moore, R. K. (2017). Is spoken language all-or-nothing? implications for future speech-based human-machine interaction. Em Jokinen, K. e Wilcock, G., editores, *Dialogues with Social Robots: Enablements, Analyses, and Evaluation*, páginas 281–291. Springer Singapore, Singapore.
- Moore, R. K., Li, H. e Liao, S. H. (2016). Progress and prospects for spoken language technology: What ordinary people think. Em *Proceedings of Interspeech 2016*, INTERSPEECH '16, páginas 3007–3011, New York, NY, USA. ACM.
- Moustakas, K., Tzovaras, D., Dybkjaer, L., Bernsen, N. e Aran, O. (2011). Using modality replacement to facilitate communication between visually and hearing-impaired people. *IEEE MultiMedia*, 18(2):26–37.

- Neto, A. T., Bittar, T. J., Fortes, R. P. M. e Felizardo, K. (2009). Developing and evaluating web multimodal interfaces - a case study with usability principles. Em *Proceedings of the 2009 ACM Symposium on Applied Computing, SAC '09*, páginas 116—120, New York, NY, USA. ACM.
- Nielsen, J. (1993). *Usability Engineering*. Academic Press, Boston.
- O'Brien, H. L. e Toms, E. G. (2008). What is User Engagement? A Conceptual Framework for Defining User Engagement with Technology. *Journal of the Association for Information Science and Technology*, 59(6):938–955.
- Park, S. Y. (2009). An analysis of the technology acceptance model in understanding university students' behavioral intention to use e-learning. *Journal of Educational Technology & Society*, 12(3):150–162.
- Peng, L. K., Ramaiah, C. K. e Foo, S. (2011). Heuristic-based user interface evaluation at nanyang technological university in singapore. *Program: electronic library and information systems*, 38(1):42–59.
- Polkosky, M. D. (2005). *Toward a Social-Cognitive Psychology of Speech Technology: Affective Responses to Speech-based E-service*. Tese de doutorado, Universidade do Sul da Flórida, Flórida, EUA.
- Porcheron, M., Fischer, J. E., Reeves, S. e Sharples, S. (2018). Voice interfaces in everyday life. Em *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, CHI '18*, página 1–12, New York, NY, USA. ACM.
- Profanter, S., Perzylo, A., Somani, N., Rickert, M. e Knoll, A. (2015). Analysis and Semantic Modeling of Modality Preferences in Industrial Human-Robot Interaction. Em *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, páginas 1812–1818, Hamburgo, Alemanha.
- Quiñones, D., Rusu, C. e Rusu, V. (2018). A Methodology to Develop Usability/User Experience Heuristics. *Computer Standards & Interfaces*, 59:109–129.
- Ren, R., Castro, J. W., Acuña, S. T. e de Lara, J. (2019). Evaluation Techniques for Chatbot Usability: A Systematic Mapping Study. *International Journal of Software Engineering and Knowledge Engineering*, 29(11):1673–1702.
- Rocha, T., Carvalho, D., Bessa, M., Reis, S. e Magalhães, L. (2017). Usability Evaluation of Navigation Tasks by People with Intellectual Disabilities: a Google and SAPO Comparative Study Regarding Different Interaction Modalities. *Universal Access in the Information Society*, 16(3):581–592.

- Roto, V., Obrist, M. e Mattila, V. V. (2009). User Experience Evaluation Methods in Academic and Industrial Contexts. Em *Proceedings of the Workshop on User Experience Evaluation Methods UXEM'09*.
- Rybarczyk, Y., Cointe, C., Gonçalves, T., Minhoto, V., Deters, J. K., Villarreal, S., Gonzalo, A. A., Baldeón, J. e Esparza, D. (2018). On the Use of Natural User Interfaces in Physical Rehabilitation: A Web-based Application for Patients with Hip Prosthesis. *Journal of Science and Technology of the Arts*, 10(2):581–592.
- Saerbeck, M., Schut, T., Bartneck, C. e Janse, M. D. (2010). Expressive robots in education: Varying the degree of social supportive behavior of a robotic tutor. Em *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '10, página 1613–1622, New York, NY, USA. Association for Computing Machinery.
- Santos, G., Rocha, A. R., Conte, T., Barcellos, M. P. e Prikladnicki, R. (2012). Strategic alignment between academy and industry: A virtuous cycle to promote innovation in technology. Em *2012 26th Brazilian Symposium on Software Engineering*, páginas 196–200, USA. IEEE.
- Schaffer, S., Schleicher, R. e Möller, S. (2015). Modeling input modality choice in mobile graphical and speech interfaces. *International Journal of Human-Computer Studies*, 75:21–34.
- Shishido, Y., Tsukagoshi, T., Yasuda, R., Vallejo, V., Tarnanas, I., Yamaguchi, T., Harada, T. e Nef, T. (2015). Adaptive Prompt System using a Ghost Shadowing Approach: a Preliminary Development. Em *Proceedings of 2015 International Conference on Virtual Rehabilitation (ICVR)*, páginas 168–169, Valência, Espanha.
- Shull, F., Carver, J. e Travassos, G. H. (2001). An empirical methodology for introducing software processes. Em *Proceedings of the 8th European Software Engineering Conference Held Jointly with 9th ACM SIGSOFT International Symposium on Foundations of Software Engineering*, ESEC/FSE-9, página 288–296, New York, NY, USA. Association for Computing Machinery.
- Spiliotopoulos, D., Stavropoulou, P. e Kouroupetroglou, G. (2009). Spoken dialogue interfaces: Integrating usability. Em A., H. e K., M., editores, *HCI and Usability for e-Inclusion*, páginas 484–499, Berlin, Heidelberg. Springer International Publishing.
- Spillane, B., Gilmartin, E., Saam, C., Su, K., Cowan, B. R., Lawless, S. e Wade, V. (2017). Introducing adele: A personalized intelligent companion. Em *Proceedings of the 1st ACM SIGCHI International Workshop on Investigating Social Interactions with Artificial Agents*, ISIAA 2017, página 43–44, New York, NY, USA. Association for Computing Machinery.
- Strauss, A. e Corbin, J. M. (2014). *Basics of Qualitative Research. Techniques and Procedures for Developing Grounded Theory*. Sage Publications.

- Su, C.-J., Chiang, C.-Y. e Huang, J.-Y. (2014). Kinect-enabled Home-based Rehabilitation System using Dynamic Time Warping and Fuzzy Logic. *Applied Soft Computing*, 22:652–666.
- Sun, H.-M. e Cheng, H.-H. (2014). The Analogical Transfer Effect of User's Experience on Usability for Gesture Control Interface. Em *Proceedings of 18th Pacific Asia Conference on Information Systems*, páginas 1–9, Chengdu, China.
- Tang, T. Y., Falzarano, M. e Morreale, P. A. (2018). Assessment of the Utility of Gesture-based Applications for the Engagement of Chinese Children with Autism. *Universal Access in the Information Society*, 17(2):275–290.
- Tchankue, P., Wesson, J. e Vogts, D. (2012). Are mobile in-car communication systems feasible? a usability study. Em *Proceedings of the South African Institute for Computer Scientists and Information Technologists Conference*, páginas 262–269, New York, NY, USA. ACM.
- Trajkova, M. e Martin-Hammond, A. (2020). “alexa is a toy”: Exploring older adults' reasons for using, limiting, and abandoning echo. Em *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, página 1–13, New York, NY, USA. ACM.
- Turunen, M., Hakulinen, J., Melto, A., Heimonen, T., Laivo, T. e Hella, J. (2009). Suxes - user experience evaluation method for spoken and multimodal interaction. Em *Proceedings of INTERSPEECH 2009*, páginas 2567–2570. INTERSPEECH.
- Uebbing-Rumke, M., Gürlük, H., Jauer, M. e Hagemann, K. and Udovic, A. (2014). Usability Evaluation of Multi-Touch-Displays for TMA Controller Working Positions. Em *Proceedings of the 4th SESAR Innovation Days*, Bélgica.
- Väätäjä, H., Koponen, T. e Roto, V. (2009). Developing practical tools for user experience evaluation: A case from mobile news journalism. Em *European Conference on Cognitive Ergonomics: Designing beyond the Product – Understanding Activity and User Experience in Ubiquitous Environments*, ECCE '09, FI-02044 VTT, FIN. VTT Technical Research Centre of Finland.
- Vardoulakis, L. P., Ring, L., Barry, B., Sidner, C. L. e Bickmore, T. (2012). Designing relational agents as long term social companions for older adults. Em Nakano, Y., Neff, M., Paiva, A. e Walker, M., editores, *Intelligent Virtual Agents*, páginas 289–302, Berlin, Heidelberg. Springer Berlin Heidelberg.
- Venkatesh, V. e Davis, F. D. (2000). A theoretical extension of the technology acceptance model: Four longitudinal field studies. *Management Science*, 46(2):186–204.
- Vetere, F., O'Hara, K., Paay, J., Ploderer, B., Harper, R. e Sellen, A. (2014). Social NUI: Social Perspectives in Natural User Interfaces. Em *Proceedings of the 2014 Companion Publication on Designing Interactive Systems*, páginas 215–218, Vancouver.

- Vosinakis, S., Koutsabasis, P., Makris, D. e Sagia, E. (2016). A Kinesthetic Approach to Digital Heritage Using Leap Motion: The Cycladic Sculpture Application. Em *Proceeding of the 8th International Conference on Games and Virtual Worlds for Serious Applications (VS-GAMES)*, páginas 1–8, Barcelona.
- Wagnier, P., Benveniste, S., Jouvelot, P. e Rigaud, A. (2018). Usability assessment of interaction management support in louise, an eca-based user interface for elders with cognitive impairment. *Technology and Disability*, 30(3):105–126.
- Wenceslao, S. J. M. C. e Estuar, M. R. J. E. (2019). Using ctakes to build a simple speech transcriber plugin for an emr. Em *Proceedings of the Third International Conference on Medical and Health Informatics 2019, ICMHI 2019*, páginas 78—86, New York, NY, USA. ACM.
- Wulf, L., Garschall, M., Himmelsbach, J. e Tscheligi, M. (2014). Hands free - care free: Elderly people taking advantage of speech-only interaction. Em *Proceedings of the 8th Nordic Conference on Human-Computer Interaction: Fun, Fast, Foundational, NordiCHI '14*, página 203–206, New York, NY, USA. ACM.
- Zhu, D., Gedeon, T. e Taylor, K. (2011). "Moving to the centre": A Gaze-Driven Remote Camera Control for Teleoperation. *Interacting with Computers*, 23:85–95.

APÊNDICE A – LISTA DOS ARTIGOS EXTRAÍDOS NO PRIMEIRO MSL

Tabela A.1: Lista dos artigos extraídos no primeiro MSL

N	Autor (es). Ano.	Título
SCOPUS_003	Vallejo V., Tarnanas I., Yamaguchi T., Tsukagoshi T., Yasuda R., Müri R., Mosimann U.P., Nef T. 2016.	Usability assessment of natural user interfaces during serious games: Adjustments for dementia intervention
SCOPUS_004	Falcao C., Lemos A.C., Soares M. 2015.	Evaluation of Natural User Interface: A Usability Study Based on the Leap Motion Device
SCOPUS_005	Ismail N.A., Pang Y.Y. 2015.	A multimodal interaction for map navigation and evaluation study of its usability
SCOPUS_013	Macaranas A., Antle A.N., Riecke B.E. 2015.	What is Intuitive Interaction? Balancing Users' Performance and Satisfaction with Natural User Interfaces
SCOPUS_014	Carvalho D., Bessa M., Magalhães L., Carrapatoso E. 2016.	Age group differences in performance using diverse input modalities: Insertion task evaluation
SCOPUS_017	Schroder S., Loftfield N., Langmann B., Frank K., Reithmeier E. 2014.	Contactless operating table control based on 3D image processing
SCOPUS_020	Kawamoto A.L.S., Martins V.F., Da Silva F.S.C. 2014	Converging natural user interfaces guidelines and the design of applications for older adults
SCOPUS_023	Shishido Y., Tsukagoshi T., Yasuda R., Vallejo V., Tarnanas I., Yamaguchi T., Harada T., Nef T. 2015.	Adaptive prompt system using a ghost shadowing approach: A preliminary development
SCOPUS_026	Canbulut C. 2017.	Usability of user interfaces based on hand gestures implemented using kinect-II and Leap Motion devices

SCOPUS_027	Erazo O., Pino J.A. 2015.	Predicting task execution time on Natural User Interfaces based on touchless hand gestures
SCOPUS_028	Erazo O., Pino J.A. 2018.	Predicting user performance time for hand gesture interfaces
SCOPUS_032	Rocha T., Carvalho D., Bessa M., Reis S., Magalhães L. 2017.	Usability evaluation of navigation tasks by people with intellectual disabilities: a Google and SAPO comparative study regarding different interaction modalities
SCOPUS_033	Cordeiro D'Ornellas M., Cargnin D.J., Cervi Prado A.L. 2015.	Evaluating the Impact of Player Experience in the Design of a Serious Game for Upper Extremity Stroke Rehabilitation.
SCOPUS_038	Kurschl W., Augstein M., Burger T., Pointner C. 2014.	User modeling for people with special needs
SCOPUS_039	Uebbing-Rumke M., Gürlük H., Jauer M., Hagemann K., Udovic A.	Usability evaluation of multi-touch-displays for TMA controller working positions
SCOPUS_043	Rybarczyk Y., Cointe C., Gonçalves T., Minhoto V., Deters J.K., Villarreal S., Gonzalo A.A., Baldeón J., Esparza D. 2018.	On the use of natural user interfaces in physical rehabilitation: A web-based application for patients with hip prosthesis
SCOPUS_044	Milani F., Rovadosky D.N., De Ávila Mendes T., Andreis J.H.D., De Marchi A.C.B., Rieder R. 2016.	Usability evaluation of menus in a gesture-based game
SCOPUS_045	Kirst D., Bulling A. 2016.	On the verge: Voluntary convergences for accurate and precise timing of gaze input
SCOPUS_053	Zhu D., Gedeon T., Taylor K.	"Moving to the centre": A gaze-driven remote camera control for teleoperation
SCOPUS_055	Cohen L., Haliyo S., Chetouani M., Régnier S. 2014.	Intention prediction approach to interact naturally with the microworld
SCOPUS_056	Sun H.-M., Cheng H.-H. 2014.	The analogical transfer effect of user's experience on usability for gesture control interface

SCOPUS_059	Economou D., Doumanis I., Argyriou L., Georgalas N. 2017.	User experience evaluation of human representation in collaborative virtual environments
SCOPUS_065	Tang T.Y., Falzarano M., Morreale P.A. 2018.	Assessment of the utility of gesture-based applications for the engagement of Chinese children with autism
SCOPUS_068	Hsiao S.-W., Lee C.-H., Yang M.-H., Chen R.-Q. 2017.	User interface based on natural interaction design for seniors
SCOPUS_072	McCaffery J., Miller A., Kennedy S., Dawson T., Allison C., Vermehren A., Lefley C., Strickland K. 2013.	Exploring Heritage through time and space supporting community reflection on the highland clearances
SCOPUS_073	Ashok V., Puzis Y., Borodin Y., Ramakrishnan I.V. 2017.	Web screen reading automation assistance using semantic abstraction
SCOPUS_079	Fiorentino M., Radkowski R., Boccaccio A., Uva A.E. 2016.	Magic mirror interface for augmented reality maintenance: An automotive case study
SCOPUS_096	Tang G., Webb P. 2018.	The Design and Evaluation of an Ergonomic Contactless Gesture Control System for Industrial Robots
SCOPUS_097	Lee J., Lee C., Kim G.J. 2017.	Vouch: multimodal touch-and-voice input for smart watches under difficult operating conditions
SCOPUS_099	Di Nuovo A., Broz F., Wang N., Belpaeme T., Cangelosi A., Jones R., Esposito R., Cavallo F., Dario P. 2018.	The multi-modal interface of Robot-Era multi-robot services tailored for the elderly
SCOPUS_102	Eckert M., Gómez-Martinho I., Meneses J., Martínez J.-F. 2017.	New approaches to exciting exergame-experiences for people with motor function impairments
SCOPUS_103	Postolache G., Carry F., Lourenço F., Ferreira D., Oliveira R., Girão P.S., Postolache O. 2017.	Serious Game for Physical Rehabilitation: Measuring the Effectiveness of Virtual and Real Training Environments
SCOPUS_112	Gustavsson P., Syberfeldt A., Brewster R., Wang L. 2017.	Human-robot Collaboration Demonstrator Combining Speech Recognition and Haptic Control

SCOPUS_114	Profanter S., Perzylo A., Somani N., Rickert M., Knoll A. 2015.	Analysis and semantic modeling of modality preferences in industrial human-robot interaction
SCOPUS_124	Mustafa, Z., Flores, J., Cotos, J.M. 2018.	Multimodal user interaction for GIS applications (MUI-GIS)
SCOPUS_130	Rodrigues, M.A.F., Serpa, Y.R., Macedo, D.V., Sousa, E.S. 2018.	A serious game to practice stretches and exercises for a correct and healthy posture
SCOPUS_132	Perez Medina, J.L., Gonzalez, M., Pilco, H.M., Beatriz Jimenes Vargas, K., Acosta-Vargas, P., Sanchez-Gordon, S., Calle-Jimenez, T., Esparza, D., Rybarczyk, Y. 2019.	Usability study of a web-based platform for home motor rehabilitation
SCOPUS_136	Muender, T., Gulani, S.A., Westendorf, L., Verish, C., Malaka, R., Shaer, O., Cooper, S. 2019.	Comparison of mouse and multi-touch for protein structure manipulation in a citizen science game interface
IEEEExplore_004	Nestorov, N., Hughes, P., Healy, N., Sheehy, N., OHare, N. 2016.	Application of Natural User Interface Devices for Touch-Free Control of Radiological Images During Surgery
IEEEExplore_009	Hsu, F., Lin, W. 2012.	Human-oriented interaction with a TOF sensor
IEEEExplore_012	Kondori, F. A., Yousefit, S., Ostovar, A., Liu, L., Li, H. 2014.	A Direct Method for 3D Hand Pose Recovery
IEEEExplore_013	Guimarães, M. D. P., Martins, V. F., Brega, J. R. F. 2012.	A software development process model for gesture-based interface
IEEEExplore_015	Kazuma, T., Yoshida, E., Yu Y., He A. 2016.	PseudoHandWriting: New Approach for Oral Presentation to have Both Advantages of Slide and Handwriting
IEEEExplore_016	Madni. T. M., Nayan, Y. B., Sulaiman S., Abro A., Tahir M. 2016.	Usability evaluation of orientation techniques for medical image analysis using tabletop system
IEEEExplore_017	Carvalho D., Bessa M., Magalhães L., Melo M., Carrapatoso E. 2016.	Age group differences in performance using distinct input modalities: A target acquisition performance evaluation

IEEEExplore_024	Vosinakis, S., Koutsabasis, P., Makris, D., Sagia, E. 2016.	A Kinesthetic Approach to Digital Heritage Using Leap Motion: The Cycladic Sculpture Application
IEEEExplore_030	Bacíková, M., Maricák, M., Vancík, M. 2015.	Usability of a domain-specific language for a gesture-driven IDE
IEEEExplore_031	Caggianese, G., Gallo, L., Pietro, G. D. 2015.	Design and Preliminary Evaluation of a Touchless Interface for Manipulating Virtual Heritage Artefacts
IEEEExplore_034	Zhao, L., Lu, X., Tao, X., Chen X. 2016.	A Kinect-Based Virtual Rehabilitation System through Gesture Recognition
IEEEExplore_043	Fabroyir, H., Teng, W., Wang, S., Tara R. Y. 2013.	MapXplorer Handy: An Immersive Map Exploration System Using Handheld Device
IEEEExplore_045	Chatzidaki, E., Xenos, M. 2017.	A case study on learning through natural ways of interaction
IEEEExplore_054	Li, A. X., Lou, X., Hansen, P., Peng, R. 2016.	On the Influence of Distance in the Interaction With Large Displays
IEEEExplore_057	Mäkelä, V., James, J., Keskinen, T., Hakulinen, J., Turunen, M. 2017.	"It's Natural to Grab and Pull": Retrieving Content from Large Displays Using Mid-Air Gestures
ScienceDirect_014	Su, C., Chiang, C., Huang, J. 2014.	Kinect-enabled home-based rehabilitation system using Dynamic Time Warping and fuzzy logic
ScienceDirect_023	Derboven, J., De Roeck, D., Verstraet, M. 2012.	Semiotic analysis of multi-touch interface design: The MuTable case study
ScienceDirect_024	Deng, S., Jiang, N., Chang, J., Guo, S., Zhang, J. J. 2017.	Understanding the impact of multi-modal interaction using gaze informed mid-air gesture control in 3D virtual objects manipulation

APÊNDICE B – LISTA DE TECNOLOGIAS DE AVALIAÇÃO RETORNADAS NO PRIMEIRO MSL

Tabela B.1: Lista de tecnologias de avaliação retornadas no primeiro MSL

System Usability Scale (SUS)	O SUS é uma tecnologia para medir a Usabilidade de produtos/serviços. É um questionário de 10 itens que pode ser respondido por meio de uma escala Likert de 5 pontos, onde 1 representa “discordo fortemente” e 5 representa “concordo fortemente”.
Análise dos resultados do estudo	A análise dos resultados do estudo consiste nos pesquisadores obterem conclusões a partir de números alcançados no experimento. Por exemplo, realizar uma análise de eficácia a partir do score alcançado pelos participantes no experimento ou ainda, analisar a eficiência por meio do tempo levado para executar uma tarefa.
Questionário criado pelos autores	São questionários desenvolvidos pelos próprios autores/pesquisadores do artigo. Geralmente são questionários específicos, onde as questões são criadas de acordo com a necessidade e contexto do estudo.
Observação direta	A observação direta consiste basicamente de um ou mais pesquisadores utilizando ou não uma estrutura de observação durante a execução do experimento como objetivo de registrar possíveis desvios de padrão. O pesquisador responsável por observar deve anotar percepções como por exemplo reações dos participantes e opiniões dadas durante o experimento para possível análise posterior.

Análise do vídeo do estudo	A análise do vídeo do experimento é uma forma de avaliação que se baseia na gravação visual de todo o experimento. Os pesquisadores gravam com uma câmera toda execução e, após o experimento, visualizam o vídeo diversas vezes para captar possíveis aspectos que considerem importante.
Entrevista criada pelos autores	São entrevistas criadas pelos próprios autores/-pesquisadores do artigo. Geralmente são entrevistas específicas, onde as questões e o roteiro são criados de acordo com a necessidade e contexto do estudo.
Think Aloud	O think aloud consiste em o participante dizer em voz alta o que ele está pensando enquanto realiza determinadas tarefas. Isto fornece aos pesquisadores uma visão dos processos cognitivos do participante. Os pesquisadores tomam notas do que os participantes dizem e fazem, especialmente observando os momentos onde encontram dificuldades.
The Touchless Hand Gesture Level Model (THGLM)	O THGLM permite estimar o tempo necessário para executar uma tarefa usando gestos. É baseado em operadores como preparo mental, tempo de respostas, preparação, retração e apontamento.
User Experience Questionnaire (UEQ)	Uma tecnologia baseada em escala que avalia a atratividade, perspicácia, eficiência, dependabilidade, estimulação e novidade. É composta de 26 itens de pares opostos.
NASA Task Load Index (NASA-TLX)	O NASA-TLX é uma ferramenta de avaliação subjetiva utilizada para classificar a carga de trabalho percebida a fim de avaliar uma tarefa ou sistema. Os aspectos avaliados por essa ferramenta são demanda mental, demanda física, demanda temporal, performance, esforço e frustração.

USE Questionnaire	O Questionário de utilidade, satisfação e facilidade de uso mede a Usabilidade subjetiva de um produto ou serviço. É um questionário de 30 itens que examina quatro dimensões da Usabilidade: utilidade, facilidade de uso, facilidade de aprendizado e satisfação.
Heurísticas de Nielsen	São 10 princípios de avaliação da Usabilidade de interfaces. Os princípios são visibilidade de qual estado o sistema está, correspondência entre o sistema e o mundo real, liberdade de controle fácil pro usuário, consistência e padrões, prevenções de erros, reconhecimento em vez de memorização, flexibilidade e eficiência de uso, estética e design minimalista, ajudar os usuários a reconhecerem, diagnosticarem e recuperarem-se de erros, ajuda e documentação.
Perceived Competence Scale (PCS)	O PCS é um questionário de 4 itens que avalia os sentimentos de competência dos participantes de um experimento. Como o PCS diz respeito a determinados comportamentos ou domínios comportamentais, ele pode ser facilmente adaptado para diversos contextos.
Web Usability Questionnaire (WUQ)	Um questionário de Usabilidade para Web contendo 20 itens que podem ser respondido por uma escala Likert.
Questionário adaptado de Kalckert e Ehrson (2014)	Um questionário simples contendo 3 itens que podem ser respondidos por uma escala Likert de 7 pontos. O objetivo deste questionário é verificar aspectos de controle do usuário.
Gesture Usability Testing Framework	Uma tecnologia que avalia a eficiência de tipos de entradas de gestos diferentes, como o Kinect e Leap Motion. Após a realização das tarefas, o framework salva o tempo e calcula a eficiência de acordo com as características de cada entrada.

Game Experience Questionnaire (GEQ)	Um questionário que avalia a experiência do usuário com algum jogo, com base em sete componentes: Imersão, Fluxo, Competência, Afeto Positivo e Negativo, Tensão e Desafio. O questionário possui 33 itens que podem ser respondidos por uma escala Likert de 5 pontos.
Capabilities of the Upper Extremity Questionnaire (CUE-Q)	É um questionário de 32 itens que avalia a dificuldade percebida para completar tarefas usando as extremidades superiores direita (15 itens), esquerda (15 itens), ou ambas (2 itens). Os itens podem ser classificados em uma escala de 5 pontos, variando entre incapacidade/dificuldade completa e nenhuma dificuldade.
Rating of Perceived Exertion (RPE)	É uma tecnologia que avalia o desempenho corporal durante a realização de uma atividade. Utiliza uma escala de 6 a 20 pontos, onde o usuário pode descrever o quão difícil a atividade foi.
Parâmetros desenvolvidos pelos autores	São definições de medidas, cálculos e métricas desenvolvidos pelos autores dos próprios estudos para avaliar o contexto de seu experimento.
General Usability Questionnaire (GUQ)	É um questionário de 7 itens usado para avaliar a conformidade com a expectativa do usuário, tolerância a falha, aptidão, controle, individualidade, descrição própria e postura de trabalho. Pode ser respondido por uma escala Likert de 5 pontos.
Paradigma do Relógio	Uma tecnologia para calcular a exatidão de uma entrada de dados e o tempo gasto. O ponteiro do relógio gira e deve ser parado pelos participantes o mais próximo possível da posição-alvo indicada pelo marcador triangular.
Questionário adaptado da ISO 9241-11 (1998)	Um questionário criado para avaliar a eficiência, eficácia e satisfação do usuário. Possui 12 itens que podem ser respondidos por uma escala Likert de 5 pontos.

Questionário adaptado de Bhuiyan e Pic-king (2011)	Um questionário desenvolvido para avaliar a carga de trabalho física, intuitividade e agradabilidade. Podem ser respondido por uma escala Likert de 5 pontos.
Questionário Ad-Hoc	Um questionário de 7 itens desenvolvido para avaliar a interação humano-robô. O questionário pode ser respondido por uma escala Likert de 5 pontos.
After-Scenario Questionnaire (ASQ) + Software Usability Measurement Inventory (SUMI)	O ASQ é uma escala de 3 perguntas usada para avaliar quão difícil um usuário percebeu uma tarefa em um teste de Usabilidade. Já o SUMI é um questionário de 50 itens que se destina a ser utilizado com uma amostra de usuários que participaram de algum experimento com um software. As respostas para os itens podem variar em “concordo”, “indeciso” e “discordo”.
Questionário adaptado de Tsai et al. (2008)	É um questionário que avalia o prazer da interação controlada pelo corpo utilizando um sensor de movimento. Possui 27 itens que podem ser respondidos por uma escala Likert de 5 pontos. Os aspectos avaliados são: prazer psíquico, prazer fisiológico, prazer social e prazer ideológico.
MEC Spatial Presence Questionnaire (MEC-SPQ)	Um questionário que avalia a atenção e o senso de presença espacial do participante. Possui 2 itens que podem ser respondidos por uma escala Likert de 4 pontos.
IBM Computer System Usability Questionnaire (CSUQ)	Um questionário de 19 itens para ser respondidos por uma escala Likert de 7 pontos. O questionário avalia a utilidade, qualidade da informação, qualidade da interface e Usabilidade geral.
User Experience Evaluation Method for Spoken and Multimodal Interaction (SU-XES)	É um método de avaliação para coletar métricas subjetivas relacionadas à UX. Ele captura tanto as expectativas dos usuários quanto as experiências dos usuários, tornando possível analisar o estado da aplicação e seus métodos de interação, e comparar resultados.

APÊNDICE C – LISTA DOS ARTIGOS EXTRAÍDOS NO SEGUNDO MSL

Tabela C.1: Lista dos artigos extraídos no segundo MSL

N	Autor (es). Ano.	Título
SCOPUS_01	Amith M., Azhu A., Cunningham R., Lin R., Savas L., Shay L., Chen Y., Gong Y., Boom J., Roberts K., e Tao C. 2019.	Early Usability Assessment of a Conversational Agent for HPV Vaccination
SCOPUS_03	Kusumaningayu F. e Ayu M. A. 2017.	A Web Accessing Tool for Blind and Visually Impaired People Using Bahasa Indonesia
SCOPUS_04	Azeta A. A., Inam I. A. e Daramola O. 2018.	A Voice-Based E-Examination Framework For Visually Impaired Students In Open And Distance Learning
SCOPUS_05	Wagnier P., Benveniste S., Jouvelot P. e Rigaud A-S. 2018.	Usability Assessment of Interaction Management Support in LOUISE, an ECA-based User Interface for Elders with Cognitive Impairment
SCOPUS_06	Kafle S. e Huenerfauth M. 2017.	Evaluating the Usability of Automatically Generated Captions for People who are Deaf or Hard of Hearing
SCOPUS_07	Gürkök H., Hakvoort G., Poel M. e Nijholt A. 2017.	Meeting the expectations from brain-computer interfaces
SCOPUS_18	Schaffer S., Schleicher R. e Möller S. 2014.	Modeling input modality choice in mobile graphical and speech interfaces
SCOPUS_19	Brusie T., Fijal T., Keller A., Lauff C., Barker K., Schwinck J., Calland J. F. e Guerlain S. 2015.	Usability Evaluation of Two Smart Glass Systems
SCOPUS_20	Coleti T.A., Morandini M. e De Lourdes Dos Santos Nunes F. 2014.	ErgoSV: An Environment to Support Usability Evaluation Using Face and Speech Recognition
SCOPUS_23	Tchankue P., Wesson J. e Vogts, D. 2012.	Are mobile in-car communication systems feasible? A usability study

SCOPUS_26	Stedmon A., Bayon V. e Griffiths G., 2011.	Expanding Interaction Potentials within Virtual Environments: Investigating the Usability of Speech and-Manual InputModes for Decoupled Interaction
SCOPUS_27	Gürkök H., Hakvoort G., Poel M. e Nijholt A. 2011.	User expectations and experiences of a speech and thought controlled computer game
SCOPUS_29	Hara S., Kitaoka N. e Takeda K. 2010.	Automatic detection of task-incompleted dialog for spoken dialog system based on dialog act N-gram
SCOPUS_31	Georgila K., Wolters M., Moore J. D. e Logie R. H. 2010.	The MATCH corpus: a corpus of older and younger users' interactions with spoken dialogue systems
SCOPUS_32	Hara S., Kitaoka N. e Takeda K. 2010.	Estimation method of user satisfaction using N-gram-based dialog history model for spoken dialog system
SCOPUS_33	Spiliotopoulos D., Stavropoulou P. e Kouroupetroglou G. 2009.	Spoken Dialogue Interfaces: Integrating Usability
SCOPUS_34	Neto A. T., Bittar T. J., Fortes R. P. M. e Felizardo, K. 2009.	Developing and evaluating web multimodal interfaces - A case study with usability principles
SCOPUS_35	Chang J. C., Lien A., Lathrop B. e Hees, H. 2009.	Usability evaluation of a Volkswagen Group in-vehicle speech system
SCOPUS_37	Anderson J. N., Davidson N., Morton H. e Jack M. A. 2008.	Language learning with interactive virtual agent scenarios and speech recognition: Lessons learned
SCOPUS_38	Fernandez Martinez F., Blazquez J., Ferreiros J., Barra R., Macias-Guarasa J. e Lucas-Cuesta J. M. 2008.	Evaluation of a Spoken Dialogue System for Controlling a Hifi Audio System
SCOPUS_39	Bernhaupt R., Palanque P., Winkler M. e Navarre D. 2007.	Usability Study of multi-modal interfaces using Eye-Tracking
SCOPUS_41	Bach P. M. e Lai J. 2006.	Usability and Learning in a Speech-Enabled Reading Tutor: A Field Study
SCOPUS_45	Toledano D. T., Pozo R. F., Trapote I. H. e Gómez L. H. 2006.	Usability evaluation of multi-modal biometric verification systems

SCOPUS_47	Davidson N., McInnes F. e Jack M. A. 2004.	Usability of dialogue design strategies for automated surname capture
SCOPUS_48	Narayanan S. e Potamianos A. 2002.	Creating Conversational Interfaces for Children
SCOPUS_50	Hone K.S. e Graham R. 2000.	Towards a tool for the Subjective Assessment of Speech System Interfaces (SASSI)
SCOPUS_52	Komine K., Hiruma N., Ishihara T., Makino E., Tsuda T., Ito T. e Isono H. 2000.	Usability Evaluation of Remote Controllers for Digital Television Receivers
IEEE_03	Miguel-Hurtado O., Blanco-Gonzalo R., Guest R. e Lunerti C. 2016.	Interaction evaluation of a mobile voice authentication system
IEEE_06	Moustakas K., Tzovaras D., Dybkjaer L., Bernsen N. O. e Aran O. 2011.	Using Modality Replacement to Facilitate Communication between Visually and Hearing-Impaired People
IEEE_07	Reis T., de Sa M. e Carrivo L. 2008.	Multimodal Artefact Manipulation: Evaluation in Real Contexts
EngVillage_06	Farinazzo V., Salvador M., Kawamoto A. L. S. e Neto J. S. O. 2010.	An empirical approach for the evaluation of voice user interfaces
EngVillage_10	Federico M. 1999.	An empirical approach for the evaluation of voice user interfaces
EngVillage_26	Carrino S., Péclat A., Mugelini E., Khaled O. A. e Ingold R. 2011.	Humans and smart environments: A novel multimodal interaction approach
EngVillage_38	Dang N. T., Tavanti M., Rankin I. e Cooper M. 2009.	A comparison of different input devices for a 3D environment
ACM_01	Huerta O., Sánchez J. A., Fuentes S. e Cervantes O. 2011.	Speak Up Your Mind: Using Speech to Capture Innovative Ideas on Interactive Surfaces
ACM_03	Tchankue P., Vogts D. e Wesson J. 2010.	Design and Evaluation of a Multimodal Interface for In-car Communication Systems
ACM_09	Wenceslao S. J. M. C. e Estuar, M. R. J. E. 2019.	Using cTAKES to Build a Simple Speech Transcriber Plugin for an EMR
ACM_10	Ferracani A., Faustino M., Giannini G. X., Landucci L. e Bimbo A. D. 2017.	Natural Experiences in Museums Through Virtual Reality and Voice Commands

ACM_13	Lumsden J., Langton N. e Kondratova I. 2008.	Evaluating the Appropriateness of Speech Input in Marine Applications: A Field Evaluation
--------	--	---

APÊNDICE D – LISTA DE TECNOLOGIAS DE AVALIAÇÃO RETORNADAS NO SEGUNDO MSL

Tabela D.1: Lista de tecnologias de avaliação retornadas no segundo MSL

Questionário criado para o estudo	São questionários desenvolvidos pelos próprios autores/pesquisadores do artigo. Geralmente são questionários específicos, onde as questões são criadas de acordo com a necessidade e contexto do estudo.
Análise da pontuação dos estudos	A análise das pontuações do estudo consiste nos pesquisadores obterem conclusões a partir de números alcançados no experimento. Por exemplo, realizar uma análise de eficácia a partir do score alcançado pelos participantes no experimento ou ainda, analisar a eficiência por meio do tempo levado para executar uma tarefa.
System Usability Scale (SUS)	O SUS é uma tecnologia para medir a Usabilidade de produtos/serviços. É um questionário de 10 itens que pode ser respondido por meio de uma escala Likert de 5 pontos, onde 1 representa “discordo fortemente” e 5 representa “concordo fortemente”.
NASA Task Load Index (NASA-TLX)	O NASA-TLX é uma ferramenta de avaliação subjetiva utilizada para classificar a carga de trabalho percebida a fim de avaliar uma tarefa ou sistema. Os aspectos avaliados por essa ferramenta são demanda mental, demanda física, demanda temporal, performance, esforço e frustração.
Entrevista criada para o estudo	São entrevistas criadas pelos próprios autores/-pesquisadores do artigo. Geralmente são entrevistas específicas, onde as questões e o roteiro são criados de acordo com a necessidade e contexto do estudo.

Observação direta	A observação direta consiste basicamente de um ou mais pesquisadores utilizando ou não uma estrutura de observação durante a execução do experimento como objetivo de registrar possíveis desvios de padrão. O pesquisador responsável por observar deve anotar percepções como por exemplo reações dos participantes e opiniões dadas durante o experimento para possível análise posterior.
UX Evaluation Method for Spoken and Multimodal Interaction (SUXES)	É um método de avaliação para coletar métricas subjetivas relacionadas à UX. Ele captura tanto as expectativas dos usuários quanto as experiências dos usuários, tornando possível analisar o estado da aplicação e seus métodos de interação, e comparar resultados.
Análise do vídeo do estudo	A análise do vídeo do experimento é uma forma de avaliação que se baseia na gravação visual de todo o experimento. Os pesquisadores gravam com uma câmera toda execução e, após o experimento, visualizam o vídeo diversas vezes para captar possíveis aspectos que considerem importante.
Questionário adaptado de Cohen et al. (2004)	Um questionário simples com 3 itens para avaliar a facilidade uso, capacidades esperadas e rapidez e eficiência. O questionário poderia ser respondido por uma escala Likert de 7 pontos.
Questionário adaptado de Iannella (2001)	Questionário de 6 itens para avaliar a funcionalidade de serviço/sistema. Este questionário pode ser respondido por uma escala Likert de 5 pontos variando entre discordo fortemente e concordo fortemente.
Questionário adaptado de Peng et al. (2011)	Questionário de 18 itens podendo ser respondido por uma escala Likert de 5 pontos. Os aspectos avaliados são: feedback, concordância entre o sistema e o mundo real, controle e liberdade do usuário, eficiência, prevenção de erro, consistência e padrão, ajuda, documentação, Usabilidade geral.

Questionário adaptado de Park (2009)	Questionário composto de 9 questões pondendo ser avaliado por uma escala Likert de 5 pontos. Os aspectos avaliados são facilidade de uso, utilidade, atitude e eficácia percebida.
Wizard of Oz	Técnica utilizada em testes de produtos para economizar recursos de produção. Ao invés de esperar que o protótipo esteja funcional para realizar testes, um operador humano funciona como computador (o mágico de oz), processando os resultados e posicionando as telas para o usuário.
Métrica Automated-Caption Evaluation (ACE)	Métrica para calcular a eficácia do reconhecimento de voz. À medida que o comprimento do texto de referência aumenta, a técnica mitiga lentamente o impacto de erros individuais - a lógica é que, como os leitores têm mais contexto (mais palavras), é mais fácil decifrar o verdadeiro significado do texto. A taxa dessa alteração é regulada de maneira não linear usando log. Porém, se o número de erros aumentar com o texto de referência, o impacto dos erros será compensado (observe a subtração de um termo $\log(n)$ no denominador).
Métrica Word Error Rate (WER)	Uma métrica para calcular a eficácia do reconhecimento de voz, baseada no número de substituições de palavra, exclusões, inserções, palavras corretas e número de palavras de referência.
Métrica Bvui	Métrica para calcular a eficiência, baseada no número de passos de uma interação por Interface Gráfica para alcançar um objetivo final e no número de passos de uma interação baseada em voz para alcançar um objetivo final.
Avaliação Heurística	Método de inspeção utilizado por arquitetos de informação e designers de interação para realizar testes de Usabilidade em interfaces de modo rápido, barato e fácil.

Questionário Input Device Usability (IDU)	Questionário que contém 15 perguntas destinadas a investigar a interação do usuário, distração, facilidade de uso, conforto do usuário, frustração, prazer, correção de erros e usabilidade geral. O questionário pode ser respondido por um escala Likert de 5 pontos.
Modelo de detecção de tarefas incompletas realizadas com a voz	Modelo criado com base em cálculos matemáticos para definir uma tarefa como completa ou incompleta.
Questionário International Telecommunication Union	Questionário de 38 itens podendo ser respondido por uma escala Likert de 5 pontos. O questionário avalia os seguintes aspectos: atingir o objetivo, comunicação com o sistema, comportamento do sistema, diálogo, avaliação pessoal e Usabilidade geral do sistema.
Ferramenta criada para o estudo	Ferramenta criada pelos autores dos trabalhos para avaliar seus estudo especificamente.
Checklist baseado nas Heurísticas de Nielsen	Adaptação das Heurísticas de Nielsen em forma de checklist, avaliando: visibilidade do status, compatibilidade do sistema com o mundo real, controle e liberdade, consistência e padronização, prevenção de erros, reconhecimento em vez de memorização, eficiência, estética, recuperação de erros e ajuda.
Questionário não-detalhado	Questionário utilizado mas com as perguntas não-detalhadas pelos autores.
Think Aloud	O think aloud consiste em o participante dizerem voz alta o que ele está pensando enquanto realiza determinadas tarefas. Isto fornece aos pesquisadores uma visão dos processos cognitivos do participante. Os pesquisadores tomam notas do que os participantes dizem e fazem, especialmente observando os momentos onde encontram dificuldades.
Attrakdiff	É uma tecnologia que avalia quão atraente o produto é em termos de usabilidade e aparência. Contém 28 itens de pares opostos que podem ser respondido por uma escala de 7 pontos.

Questionário baseado em Dutton et al. (1993)	Questionário para avaliar a atitude, composto de 14 itens que podem ser respondidos por uma escala Likert de 7 pontos.
Subjective Assessment of Speech Systems Interfaces (SASSI)	Conjunto de 34 itens respondidos por uma escala Likert que avaliam seis conjuntos de aspectos: precisão da resposta do sistema, satisfação do usuário, demanda cognitiva, aborrecimento, "habitability" e velocidade.
Análise do áudio do estudo	A análise do áudio do experimento é uma forma de avaliação que se baseia na gravação do áudio de todo o experimento. Os pesquisadores gravam toda execução e, após o experimento, ouvem a gravação diversas vezes para captar possíveis aspectos que considerem importante.
Checklist criado para o estudo	São checklists desenvolvidos pelos próprios autores/pesquisadores do artigo. Geralmente são checklists específicos, onde os itens são criados de acordo com a necessidade e contexto do estudo
Lei de Fitts	Modelo do movimento humano, que prediz o tempo necessário para mover-se rapidamente desde uma posição inicial até uma zona destino final como uma função da distância até o objetivo e o tamanho deste. Em forma de avaliação, foi utilizado para calcular o tempo de execução da posição inicial até a final.
Questionário Computer System Usability (CSUQ)	Um questionário de 19 itens para ser respondidos por uma escala Likert de 7 pontos. O questionário avalia a utilidade, qualidade da informação, qualidade da interface e Usabilidade geral.

APÊNDICE E – RESUMO 1º MSL

Tabela E.1: Respostas da SQ1, SQ2, SQ2.1, SQ2.2, SQ3 e SQ4.

SQ1: (A) Multitoque; (B) Gesto; (C) Voz; (D) Olhar. SQ2: (A) Usabilidade; (B) Experiência do Usuário; (C) Ambas. SQ2.1: (A) Teste; (B) Inspeção; (C) Investigação; (D) Modelagem Analítica; (E) Simulação. SQ2.2: (A) Estudo de Laboratório; (B) Estudo de Caso; (C) Survey; (D) Especialistas. SQ3: (A) Específica; (B) Genérica. SQ4: (A) Sim; (B) Não. - : Não aplicável.

Ref.	Tecnologia	SQ1				SQ2			SQ2.1					SQ2.2				SQ3		SQ4	
		A	B	C	D	A	B	C	A	B	C	D	E	A	B	C	D	A	B	A	B
Scopus_03	T01_1		x			x					x			-	-	-	-		x	x	
	T01_2		x			x			x					-	-	-	-		x	x	
Scopus_04	T02_1		x			x			x					-	-	-	-		x	x	
	T02_2		x			x					x			-	-	-	-		x	x	
Scopus_05	T03_01		x			x			x					-	-	-	-		x	x	
	T03_02		x			x					x			-	-	-	-		x	x	
	T04_1		x			x					x			-	-	-	-		x	x	
Scopus_13	T04_2		x			x					x			-	-	-	-		x	x	
	T04_3		x			x					x			-	-	-	-	x			x
	T04_4		x			x			x					-	-	-	-		x	x	
Scopus_14	T05_01		x			x			x					-	-	-	-		x	x	
	T05_02		x			x					x			-	-	-	-	x			x
Scopus_17	T06_01		x			x			x					-	-	-	-		x	x	
	T06_02		x			x					x			-	-	-	-		x	x	
Scopus_20	T07		x			x					x			-	-	-	-		x	x	
Scopus_23	T08		x			x					x			-	-	-	-	x		x	
Scopus_26	T09		x			x			x					-	-	-	-	x		x	
Scopus_27	T10		x			x			x					-	-	-	-	x			x
Scopus_28	T11		x			x			x					-	-	-	-	x		x	
Scopus_32	T12_1				x	x			x					-	-	-	-		x	x	
	T12_2				x	x			x					-	-	-	-		x	x	
Scopus_33	T13_1		x					x	-	-	-	-	-	x					x	x	
	T13_2		x					x	-	-	-	-	-	x					x	x	
	T13_3		x					x	-	-	-	-	-	x					x	x	
Scopus_38	T14		x			x					x			-	-	-	-	x			x
Scopus_39	T15_01	x				x					x			-	-	-	-		x	x	
	T15_02	x						x	-	-	-	-	-	x					x	x	
	T15_03	x				x					x			-	-	-	-		x	x	
	T15_04	x				x					x			-	-	-	-		x	x	
	T15_05	x				x			x					-	-	-	-		x	x	
Scopus_43	T16		x			x					x			-	-	-	-		x	x	
Scopus_44	T17_01		x			x			x					-	-	-	-		x	x	
	T17_02		x			x					x			-	-	-	-	x			x
Scopus_45	T18_01				x	x			x					-	-	-	-		x	x	
	T18_02				x	x					x			-	-	-	-		x	x	
	T19_01				x	x			x					-	-	-	-		x	x	
Scopus_53	T19_02				x	x					x			-	-	-	-	x			x
	T19_03				x	x					x			-	-	-	-	x			x
Scopus_55	T20		x			x					x			-	-	-	-		x	x	
Scopus_56	T21		x			x					x			-	-	-	-		x	x	
Scopus_59	T22		x					x	-	-	-	-	-	x					x	x	
Scopus_65	T23_01		x			x					x			-	-	-	-	x			x
	T23_02		x			x			x					-	-	-	-		x	x	
Scopus_68	T24		x			x			x					-	-	-	-		x	x	
Scopus_72	T25_01		x			x					x			-	-	-	-		x	x	
	T25_02		x					x	-	-	-	-	-	x				x			x
	T25_03		x					x	-	-	-	-	-	x				x			

Tabela E.2: Continuação da Tabela E.1

Ref.	Tecnologia	SQ1				SQ2			SQ2.1					SQ2.2				SQ3		SQ4	
		A	B	C	D	A	B	C	A	B	C	D	E	A	B	C	D	A	B	A	B
Scopus_73	T26_01			x		x					x			-	-	-	-		x	x	
	T26_02			x		x					x			-	-	-	-	x			x
	T26_03			x		x			x					-	-	-	-		x	x	
Scopus_79	T27		x			x			x					-	-	-	-		x	x	
Scopus_96	T28_01		x			x					x			-	-	-	-		x	x	
	T28_02		x			x					x			-	-	-	-	x			x
Scopus_97	T29			x		x					x			-	-	-	-	x			x
Scopus_99	T30_01			x		x					x			-	-	-	-		x	x	
	T30_02			x		x					x			-	-	-	-		x	x	
Scopus_102	T31_1		x			x					x			-	-	-	-	x			x
	T31_2		x			x			x					-	-	-	-		x	x	
Scopus_103	T32		x			x			x					-	-	-	-		x	x	
Scopus_112	T33_01			x		x					x			-	-	-	-		x	x	
	T33_02			x		x					x			-	-	-	-	x			x
	T33_03			x		x			x					-	-	-	-		x		x
Scopus_114	T34_01		x	x		x					x			-	-	-	-	x			x
	T34_02		x	x		x					x			-	-	-	-		x		x
	T34_03		x	x		x					x			-	-	-	-	x		x	
Scopus_124	T35		x	x		x					x			-	-	-	-		x	x	
Scopus_130	T36_1		x			x			x					-	-	-	-		x	x	
	T36_2		x			x					x			-	-	-	-	x			x
	T36_3		x				x		-	-	-	-	-	x					x	x	
Scopus_132	T37_1		x			x			x					-	-	-	-		x	x	
	T37_2		x			x					x			-	-	-	-		x	x	
	T38_1	x				x			x					-	-	-	-		x	x	
Scopus_136	T38_2	x				x					x			-	-	-	-		x	x	
	T38_3	x				x					x			-	-	-	-		x	x	
	T38_4	x				x					x			-	-	-	-		x	x	
IEEEExplore_4	T39_01		x			x					x			-	-	-	-		x	x	
	T39_02		x			x					x			-	-	-	-	x			x
	T39_03		x			x			x					-	-	-	-		x	x	
IEEEExplore_9	T40		x			x					x			-	-	-	-	x		x	
IEEEExplore_12	T41_01		x			x					x			-	-	-	-		x		x
	T41_02		x			x			x					-	-	-	-		x	x	
IEEEExplore_13	T42_01		x			x				x				-	-	-	-		x	x	
	T42_02		x			x			x					-	-	-	-		x	x	
IEEEExplore_15	T43		x			x			x					-	-	-	-		x	x	
IEEEExplore_16	T44	x				x					x			-	-	-	-		x	x	
IEEEExplore_17	T45_1		x			x					x			-	-	-	-	x			x
	T45_2		x			x			x					-	-	-	-		x	x	
IEEEExplore_24	T46_01		x				x		-	-	-	-	-	x					x	x	
	T46_02		x				x		-	-	-	-	-	x					x	x	
	T46_03		x				x		-	-	-	-	-	x					x		x
IEEEExplore_30	T47	x				x			x					-	-	-	-		x	x	
IEEEExplore_31	T48_01		x			x					x			-	-	-	-		x	x	
	T48_02		x			x					x			-	-	-	-		x	x	
	T48_03		x				x		-	-	-	-	-	x					x	x	
IEEEExplore_34	T49		x			x			x					-	-	-	-		x	x	
IEEEExplore_43	T50_01	x				x			x					-	-	-	-		x	x	
	T50_02	x				x					x			-	-	-	-	x			x
IEEEExplore_45	T51_01		x	x		x			x					-	-	-	-		x	x	
	T51_02		x	x			x		-	-	-	-	-	x				x			x
	T52_01		x			x			x					-	-	-	-		x	x	
IEEEExplore_54	T52_02		x			x					x			-	-	-	-		x	x	
	T52_03		x				x		-	-	-	-	-	x				x			x
	T53_01		x				x		-	-	-	-	-	x					x		
IEEEExplore_57	T53_02		x				x		-	-	-	-	-	x				x			x
	T54_01		x			x					x			-	-	-	-	x			x
SciDirect_14	T54_02		x			x			x					-	-	-	-		x	x	
SciDirect_23	T55	x				x			x					-	-	-	-		x	x	
SciDirect_24	T56_01		x		x	x			x					-	-	-	-		x	x	
	T56_02		x		x	x					x			-	-	-	-		x	x	
	T56_03		x		x	x					x			-	-	-	-	x			x

Tabela E.4: Continuação da Tabela E.3.

[illegible]

Tabela E.5: Respostas da SQ5.2 e SQ5.3.

(A) Academia; (B) Indústria; (C) Laboratório; (D) Misto. SQ5.3: (A) Quantitativo; (B) Qualitativo; (C) Ambos. - : Não aplicável

Ref.	Avaliação	SQ5.2				SQ5.3		
		A	B	C	D	A	B	C
Scopus_3	E01_1			x		x		
	E01_2			x		x		
Scopus_4	E02			x				x
Scopus_5	E03			x		x		
Scopus_13	E04			x				x
Scopus_14	E05			x				x
Scopus_17	E06			x				x
Scopus_20	E07			x				x
Scopus_23	E08_1				x	x		
	E08_2				x	x		
Scopus_26	E09			x		x		
	E10_1			x		x		
Scopus_27	E10_2			x		x		
	E10_3			x		x		
	E10_4			x		x		
Scopus_28	E11			x		x		
Scopus_32	E12			x		x		
Scopus_33	E13			x		x		
Scopus_38	E14			x				x
Scopus_39	E15			x				x
Scopus_43	E16			x		x		
Scopus_44	E17			x		x		
Scopus_45	E18			x		x		
Scopus_53	E19	x						x
Scopus_55	E20			x				x
Scopus_56	E21	-	-	-	-	-	-	-
Scopus_59	E22			x			x	
	E23_1			x			x	
Scopus_65	E23_2			x			x	
	E23_3			x			x	
Scopus_68	E24			x				x
Scopus_72	E25			x				x
Scopus_73	E26			x				x
Scopus_79	E27_1			x		x		
	E27_2			x			x	
Scopus_96	E28			x		x		
Scopus_97	E29			x		x		
Scopus_99	E30			x		x		
Scopus_102	E31			x		x		
Scopus_103	E32			x		x		
Scopus_112	E33			x		x		
Scopus_114	E34			x		x		
Scopus_124	E35			x		x		
Scopus_130	E36			x				x
Scopus_132	E37			x		x		
Scopus_136	E38			x		x		
IEEEExplore_4	E39_1			x		x		
	E39_2			x				x
IEEEExplore_9	E40			x		x		
IEEEExplore_12	E41			x		x		
IEEEExplore_13	E42			x			x	
IEEEExplore_15	E43			x			x	
IEEEExplore_16	E44			x		x		
IEEEExplore_17	E45			x				x
IEEEExplore_24	E46			x				x
IEEEExplore_30	E47			x			x	
IEEEExplore_31	E48			x		x		
IEEEExplore_34	E49			x		x		
IEEEExplore_43	E50			x		x		
IEEEExplore_45	E51			x				x
IEEEExplore_54	E52			x				x
IEEEExplore_57	E53			x				x
SciDirect_14	E54			x		x		
SciDirect_23	E55			x			x	
SciDirect_24	E56			x		x		

APÊNDICE F – RESUMO 2º MSL

Legenda para as Tabelas F.3, F.4 e F.5 descritas abaixo. Aspectos de Usabilidade. SQ2:

(a) Eficiência; (b) Satisfação do usuário; (c) Eficácia; (d) Facilidade de uso; (e) Performance; (f) Frustração; (g) Demanda mental; (h) Taxa de erro; (i) Dificuldade; (j) Uso do serviço; (k) Demanda Física; (l) Demanda Temporal; (m) Esforço; (n) Prazer; (o) Precisão da Resposta; (p) Usabilidade Geral; (q) Interatividade; (r) Heurísticas de Nielsen; (s) Distração; (t) Correção de erro; (u) Comportamento do sistema; (v) Diálogo; (w) Engajamento; (x) Naturalidade; (y) Controle da Interação; (z) Preferência; (aa) Funcionalidade; (ab) Memorabilidade; (ac) Aprendabilidade; (ad) Facilidade de Aprendizado; (ae) Capacidades esperadas; (af) Rapidez; (ag) Tempo de resposta; (ah) Repetitividade; (ai) Recursos visuais; (aj) Humanização do sistema; (ak) Habilidade de executar as tarefas; (al) Quantidade de Informação; (am) Acessibilidade; (an) Clareza do sistema; (ao) Funcionalidade do serviço; (ap) Flexibilidade; (aq) Agradabilidade; (ar) Aparência; (as) Clareza para falar; (at) Facilidade de instrução; (au) Ritmo da conversa; (av) Conforto de uso; (aw) Atingir o objetivo; (ax) Comunicação com o sistema; (ay) Especialidade do usuário; (az) Avaliação pessoal; (ba) Entendimento; (bb) Impressão; (bc) Utilidade; (bd) Facilidade da tarefa; (be) Tolerância de erro; (bf) Iniciativa de diálogo; (bg) Robustez; (bh) Uso futuro; (bi) Motivação; (bj) Qualidade; (bk) Conclusão da Tarefa; (bl) Aspectos Pragmáticos; (bm) Reações Subjetivas; (bn) Atitude; (bo) Simpatia; (bp) Aborrecimento; (bq) Habitabilidade; (br) Velocidade; (bs) Design; (bt) Singularidade; (bu) Problema de Usabilidade; (bv) Aceitação; (bw) Feedback Adequado; (bx) Diversidade de Usuário; (by) Percepção; (bz) Esforços de Memorização; (ca) Sentenças de Saída Adequadas; (cb) Qualidade da Saída de Voz; (cc) Reconhecimento de Entrada Adequado; (cd) Início do Diálogo Adequado; (ce) Instruções; (cf) Estrutura do Diálogo Natural; (cg) Suficiência de Orientação da Interface; (ch) Ferramenta de Ajuda; (ci) Prevenção de Erro; (cj) Manipulação de Erro; (ck) Simplicidade.

Tabela F.1: Respostas da SQ1, SQ3, SQ4, SQ7 e SQ11.

SQ1: (A) Usabilidade; (B) Experiência do Usuário; (C) Ambas. SQ3: (A) Específica; (B) Genérica. SQ4: (A) Criada pelo estudo. (B) Existente. SQ7: (A) Quantitativa; (B) Qualitativa; (C) Ambas. SQ11: (A) Sim; (B) Não.

Ref.	Tecnologia	SQ1			SQ3		SQ4		SQ7			SQ11	
		A	B	C	A	B	A	B	A	B	C	A	B
Scopus_1	P001_T1	x			x			x	x				x
	P001_T2	x			x		x			x			x
	P002_T1	x			x		x			x			x
Scopus_3	P002_T2	x			x			x	x				x
	P002_T3	x			x			x	x				x
	P002_T4		x		x			x	x				x
	P002_T5		x		x			x		x			x
Scopus_4	P003_T1	x			x		x		x				x
Scopus_5	P004_T1	x			x		x		x				x
Scopus_6	P005_T1	x			x		x		x			x	
	P005_T2	x			x			x	x			x	
Scopus_7	P006_T1		x			x		x	x				x
Scopus_18	P007_T1	x			x		x		x			x	
	P007_T2	x				x		x	x				x
Scopus_19	P008_T1	x				x		x		x			x
Scopus_20	P009_T1	x				x		x	x				x
	P010_T1	x				x		x	x				x
Scopus_23	P010_T2	x			x		x		x				x
	P010_T3	x				x		x	x				x
	P011_T1	x				x		x			x		x
Scopus_26	P011_T2		x		x		x		x				x
	P011_T3	x			x		x		x				x
Scopus_27	P012_T1		x			x		x	x				x
Scopus_29	P013_T1	x			x		x		x			x	
Scopus_31	P014_T1	x			x			x	x				x
Scopus_32	P015_T1	x				x		x			x		x
Scopus_33	P016_T1	x			x		x		x				x
Scopus_34	P017_T1	x				x		x		x			x
	P018_T1	x				x		x		x			x
Scopus_35	P018_T2	x				x		x	x				x
	P018_T3	x				x		x		x			x
	P019_T1	x			x			x	x				x
Scopus_37	P019_T2	x				x		x	x				x
	P019_T3	x			x		x		x				x
	P020_T1	x			x			x	x				x
Scopus_38	P020_T2	x			x		x		x				x
	P020_T3	x			x		x		x				x
Scopus_39	P021_T1	x				x		x	x				x
	P021_T2	x				x		x		x			x
	P021_T3	x				x		x	x				x
	P021_T4	x				x		x	x				x
Scopus_41	P022_T1	x				x		x		x			x
	P022_T2	x				x		x	x				x
	P022_T3	x				x		x			x		x
Scopus_45	P023_T1	x				x		x	x				x
	P023_T2	x			x		x		x				x
Scopus_47	P024_T1	x				x		x	x				x
	P024_T2	x				x		x	x				x
Scopus_48	P024_T3	x				x		x					x
	P025_T1		x		x		x				x		x

Tabela F.2: Continuação Tabela F.1

Ref.	Tecnologia	SQ1			SQ3		SQ4		SQ7			SQ11	
		A	B	C	A	B	A	B	A	B	C	A	B
Scopus_50	P026_T1	x			x		x		x			x	
	P027_T1	x				x		x	x				x
Scopus_52	P027_T2	x			x		x		x				x
	P027_T3	x			x		x		x				x
IEEEExplore_3	P028_T1	x				x		x	x				x
	P028_T2	x			x		x		x				x
	P029_T1			x	x		x			x			x
IEEEExplore_6	P029_T2		x			x		x	x				x
	P029_T3		x			x		x		x			x
	P029_T4	x				x		x	x				x
IEEEExplore_7	P030_T1	x			x		x				x		x
	P030_T1	x				x		x	x				x
EngVillage_6	P031_T1	x			x		x		x				x
EngVillage_10	P032_T1	x				x		x	x				x
	P032_T2	x			x		x		x				x
EngVillage_26	P033_T1	x				x		x	x				x
	P033_T2	x				x		x	x				x
	P033_T3	x				x		x	x				x
EngVillage_38	P034_T1	x				x		x	x				x
	P034_T2	x			x		x		x				x
ACM_001	P035_T1	x			x		x				x		x
ACM_3	P036_T1	x				x		x	x				x
	P036_T2	x				x		x	x				x
	P036_T3	x				x		x			x		x
ACM_9	P037_T1	x				x		x	x				x
ACM_10	P038_T1	x				x		x	x				x
ACM_13	P039_T1	x				x		x	x				x
	P039_T2	x				x		x	x				x

Tabela F3: Aspectos de usabilidade avaliados. Descrição de cada um acima.

[illegible]

Tabela F.4: Continuação da Tabela F.3.

[illegible]

Tabela F.5: Continuação da Tabela F.4

Ref.	Tecnologia	cf	cg	ch	ci	cj	ck
Scopus_1	P001_T1						
	P001_T2						
Scopus_3	P002_T1						
	P002_T2						
	P002_T3						
Scopus_4	P003_T1						
Scopus_5	P004_T1						
Scopus_6	P005_T1						
	P005_T2						
Scopus_18	P007_T1						
	P007_T2						
Scopus_19	P008_T1						
Scopus_20	P009_T1						
	P010_T1						
Scopus_23	P010_T2						
	P010_T3						
Scopus_26	P011_T1						
	P011_T3						
Scopus_29	P013_T1						
Scopus_31	P014_T1						
Scopus_32	P015_T1						
Scopus_33	P016_T1						
Scopus_34	P017_T1						
	P018_T1						
Scopus_35	P018_T2						
	P018_T3						
	P019_T1						
Scopus_37	P019_T2						
	P019_T3						
Scopus_38	P020_T1						
	P020_T2						
	P020_T3						
	P021_T1						
Scopus_39	P021_T2						
	P021_T3						
	P021_T4						
	P022_T1						
Scopus_41	P022_T2						
	P022_T3						
Scopus_45	P023_T1						
	P023_T2						
	P024_T1						
Scopus_47	P024_T2						
	P024_T3						
Scopus_50	P026_T1						
	P027_T1						
Scopus_52	P027_T2						
	P027_T3						
IEEEExplore_3	P028_T1						
	P028_T2						
IEEEExplore_6	P029_T1						
	P029_T4						
IEEEExplore_7	P030_T1						
EngVillage_6	P030_T1	x					
	P031_T1						
EngVillage_10	P032_T1						
	P032_T2						
	P033_T1						
EngVillage_26	P033_T2						
	P033_T3						
EngVillage_38	P034_T1						
	P034_T2						
ACM_1	P035_T1						
	P036_T1						
ACM_3	P036_T2						
	P036_T3						
ACM_9	P037_T1						
ACM_10	P038_T1						
	P039_T1						
ACM_13	P039_T2						x

Tabela F.7: Resposta da SQ5.

(a) Escala Likert; (b) Tempo gasto nas tarefas; (c) Reconhecimento de fala automático; (d) Questões abertas; (e) Taxa de erro; (f) Observação; (g) Completude das tarefas; (h) Arquivo de log; (i) Gravação do áudio; (j) Escolha da preferência; (k) Gravação de vídeo; (l) Sim ou não; (m) Percepção após uso; (n) Taxa de acerto; (o) Checklist; (p) Número de distrações; (q) Sim, não ou não aplicável.

[illegible]

Tabela F.9: Respostas da SQ10.

(a) Cegos/deficientes visuais; (b) Idosos; (c) Surdos/deficientes auditivos; (d) Pessoas com Alzheimer; (e) Avaliadores de sistema; (f) Estudantes; (g) Operadores de satélite; (h) Crianças; (i) Funcionários de escritório; (j) Deficientes motores.

Ref.	Sistema	a	b	c	d	e	f	g	h	i	j
Scopus_3	S002	x									
Scopus_4	S003	x									
Scopus_5	S004		x		x						
Scopus_6	S005			x							
Scopus_20	S009					x					
Scopus_37	S019						x				
Scopus_39	S021							x			
Scopus_48	S025								x		
Scopus_52	S027		x								
IEEEExplore_6	S029	x		x							
EngVillage_10	S032									x	
ACM_10	S038										x

Tabela F.10: Respostas da SQ12.

Ref.	Tecnologia	Estudo de Observação	Estudo de Viabilidade
Scopus_6	P005_T1		x
Scopus_6	P005_T2		x
Scopus_18	P007_T1	x	
Scopus_29	P013_T1		x
Scopus_50	P026_T1		x